

Psychological Bulletin

CONTENTS

- A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1946-1957..... IRVING JANIS, DAVID FOR, JANE DAVIS, AND MARLIN BRENNER 337
- Indicators of Perception: I. Subliminal Perception, Subception, Unconscious Perception: An Analysis in Terms of Psychophysical Indicator Methodology..... ISRAEL GOLDAMOND 373
- Spontaneous Association Behavior..... WILLIAM N. DEMER AND HARRY FOWLER 412
- The Continuity of Abnormal and Normal Behavior... H. J. EYSENCK 429
- Discontinuity and Correlation: A Reply to Eysenck..... JOHN S. PHARSON AND IRENE B. KLEY 433

This is the last issue of Volume 55

Volume contents and this page
appear herein.

Published bimonthly by the
American Psychological Association

VOL. 55, NO. 5

NOVEMBER, 1958

WAYNE DENNIS, Editor
Brooklyn College

Consulting Editors

LAUNOR F. CARTER
RAND Corporation
Santa Monica, California
EDWARD GIRDEN
Brooklyn College
VICTOR C. RAIMY
University of Colorado

ROBERT L. THORNDIKE
Teachers College, Columbia University
BENTON J. UNDERWOOD
Northwestern University
S. RAINS WALLACE
Life Insurance Agency
Management Association

ARTHUR C. HOFFMAN, Managing Editor

HELEN ORR, Promotion Manager
SARAH WOMACK, Editorial Assistant

The Psychological Bulletin contains evaluative reviews of research literature and articles on research methodology in psychology. This JOURNAL does not publish reports of original research or original theoretical articles.

Manuscripts should be sent to the Editor-elect, Harry Helson, Department of Psychology, University of Texas, Austin, Texas.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 50-60), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Reprints. Fifty free offprints are given to contributors of articles and notes. Authors of early publication articles receive no gratis offprints.

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: \$8.00 (Foreign \$8.50). Single copies, \$1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage under Government of Sec. 3640 Post. (D) provided for in Section 558, act of February 26, 1949, authorized August 6, 1949. Second-class postage paid at Menasha, Wisconsin and Washington, D.C. Printed in U.S.A.

Copyright, 1958, by The American Psychological Association, Inc.

Psychological Bulletin

A SURVEY OF STUDIES CONTRASTING THE QUALITY OF GROUP PERFORMANCE AND INDIVIDUAL PERFORMANCE, 1920-1957¹

IRVING LORGE, DAVID FOX, JOEL DAVITZ, AND MARLIN BRENNER
Teachers College, Columbia University

This is an analysis of studies done in the years 1920-1957 which contrast the quality of performance by individuals and by groups in diverse situations. A number of studies are included which add to our understanding of this aspect of human behavior. However, an unpublished review by Lorge et al. (37) prepared in 1953 served as an important source for this presentation. In fact, some of the organization of that report is carried into this study. The existence of this report is due to a literature search made in connection with research into group performance and group process in problem solving.

It is important to focus on basic concepts such as "group," "task," and "criterion." These terms have been applied in so many different senses and in so many different situations that clarification and differentiation are necessary for the interpretation of the research results.

The most ambiguous term seems to be that of "group," which not only is recognized in a variety of senses by lexicographers, but also is used with a wide range of meanings by social psychologists. The lexicog-

rapher considers a group as: (a) an assemblage of persons in physical proximity considered as a collective unity, e.g., a group by aggregation; and (b) a unity of a number of persons classed together because of any kind of common relation, whether of organization or of commitment. The social psychologist recognizes three kinds of groups: (a) an assemblage of persons in a physical environment; (b) an association of persons with some form of social, political, or managerial organization; and (c) a collective unity of members subscribing to a common symbol or loyalty. Sapir, in the *Encyclopedia of the Social Sciences*, distinguishes three classes of groups: (a) persons at a football game or in a train; (b) organizationally defined, as having some mutuality of purpose, e.g., employees in a factory or pupils in a classroom; and (c) symbolically defined, as serving some well-recognized function or functions, e.g., family, military staff, or executive cabinet.

A military staff or an executive cabinet, however, achieves its collective unity as a consequence of having interacted with one another over a considerable period of time, so that they have developed a tradition of working together for mutual and common purposes. This viewpoint allows one to think of the group as

¹ Work sponsored by the Office of Naval Research under Contract Nonr 266(43) and extending research which was supported in part by the United States Air Force under Contract No. AF 33(038)-28792.

continuously emergent—the longer its members work together, the greater the possibility of developing a more cohesive and more cooperative team. Group cohesiveness, moreover, may be one of the resultants of interaction among a team's members that leads to the development of a group or team "tradition." The social psychologist tends to *think* of the "group" as having a "tradition," i.e., a cooperative association of individuals whose members have progressed through the states of coming together in physical proximity, of organizing for common goals, and of accepting commitment for the group's purposes. The members of a traditioned group will have assayed each other as resources and as personalities, will have established channels of communication, and will have achieved mutual reinforcement for the common goal. The traditioned group, therefore, is a functioning unity—functioning for a real and genuine goal. While the world's work is accomplished by many traditioned groups or teams or staffs, such traditioned groups have not been studied extensively, primarily because of the difficulty of access and their unwillingness to have others observe their processes.

Methodologically, it is important for social psychology to develop an understanding of the changing dynamics of the *emerging* groups. To do this, social psychologists have usually worked with *ad hoc* groups, i.e., some experimenter has assembled several individuals to work together mutually and cooperatively on some specific and externally assigned task. An *ad hoc* group, therefore, may represent one end of a continuum of "group" which extends from the just-assembled *ad hoc*, to the well-

established, traditioned group. *Ad hoc* groups, necessarily, will vary in the extent of cohesion that they achieve, as well as in the acceptance of the mutuality of purposes. Each externally designated *ad hoc* group, therefore, in some more or less tentative way, must organize, test each other's resources, accept the task goal, muster its resources to reach that goal, and then accomplish its end. Such experimental *ad hoc* groups usually cease to exist when the experimenter's purposes have been achieved. The research use of the *ad hoc* groups is exemplified in the experiments of Watson (70) and of Shaw (55). They each selected college students at random from the same class to form a group for the experimenter's purposes, and then, only for the duration of the experiment.

A common and dangerous practice is to generalize the principles valid for *ad hoc* groups to traditioned groups. The *ad hoc* group is treated as a microscopic model of the traditioned group. This might be true, but has not been experimentally validated. It is equally possible that *ad hoc* and traditioned groups behave in accordance with their individual principles.

The continuum, therefore, of *ad hoc* to traditioned groups constitutes an ambiguous and complex semantic range for interacting, face-to-face groups who *deliberate* to solve problems or produce joint products. In sharp contrast to the continuum of *ad hoc* to traditioned groups which do interact among its members, are the so-called groups whose individuals do not overtly interact with one another. Rather, these are groups only because their constituent units are in physical proximity. In social psychology, groups by physical proxim-

ity have been utilized primarily in research involving an individual's performance in a sociophysical setting of other individuals. Usually, the experiments have been studies of "social facilitation," designed to appraise the psychological consequences on the individual working in a mass or among one's fellows. The sociophysical setting has been termed by some psychologists as the "climatized group" but it must be recognized that the group is a group only in the sense of having members in physical proximity.

Three variations in "climatized group" are reported in the research literature. Of these, the type nearest to the real group provides for group discussion of a problem followed by individual judgments or estimations. Such a "climatized group" has interaction among individuals but no measure of *group consensus*. The jury experiments of Bechtereve and Lange (7) and of Burr (9) illustrate the pattern, e.g., the credibility of the testimony of witnesses is discussed by the "jury," followed by judgment by each individual on the issue.

The second variety of "climatized group" does not provide for discussion, but rather is a sequel to individual evaluation of group judgment or is an evaluation made by some open form of voting, like a show of hands. Gurnee's (23) 1937 experiment employs such a "climatized group": college students first took a true-false examination as individuals, and then repeated the same examination as a group. Group choice was determined by a show of hands. While "visual" interaction by observations of other members' voting behavior was evident, it certainly was not the overt verbal interaction of deliberation.

The third variety of "climatized" group has neither interaction nor consensus. In such a "climatized group," the individual works alone at his task in the presence of other people, as, for example, in the social facilitation experiments of Allport (1) and of Dashiell (15), in which individuals either worked in isolation or in the presence of others, but without any interaction.

The research literature frequently refers to a type of group which is really not a group at all—rather, it is a consequent of statistical computations, i.e., averaging of the products of independent and noninteracting individuals. The "statisticized" group, for instance, was used in the 1924 study of Kate Gordon (21) in which college students *as individuals* judged weights. These individual judgments, then, were averaged to form "groups" of 5, or of 10, or of 20, or of 50. Since such a statistical "group" neither meets nor interacts, it does not function as a psychological entity. It is of dubious semantic advantage to designate the consequence of such statistical averaging or aggregating a "group" product. Experiments with the "statisticizing" technique may be more appropriately considered as evidence about the reliability of measurement (of one judge versus several judges) rather than about group dynamics. Basically, the "statisticized" group appraises aggregation, not interaction.

Another so-called group is the "concocted" group. It, too, neither meets nor interacts. In the "concocted" group the unique elements of *each individual's* products are combined to form a so-called *group* product. One form of "concocted" group is that in which each individual's products are *summed* to form the so-

called *group* product. A second form is represented in Marquart's experiment of 1955 (42). Individual's products are combined so that one "solution" is assigned a fictitious group if at least one of the members *working individually* arrives at the correct solution. A "no solution" is assigned the group if none of the "group" members arrive at the correct solution individually.

In the experimental literature, the earliest studies were of the "statisticized" group, followed subsequently and in succession by the "climatized," the "concocted," the *ad hoc*, and most recently the "traditioned" group. Since the "traditioned" group is most like real life groups, the development may be considered to have moved along a continuum from artificial ("statisticized") to real ("traditioned").

The varieties of groups may be broadly classified, then, as follows:

1. Interacting, face-to-face group, i.e., involving group meeting and discussion:
 - a. with a tradition of working together (traditioned)
 - b. with no tradition of working together (*ad hoc*)
2. Noninteracting face-to-face group, i.e., involving physical meeting, but no discussion:
 - a. with a sequel appraisal of group opinion (climatized)
 - b. with a sequel appraisal of individual opinion (social climatized)
3. Noninteracting non-face-to-face group, i.e., involving no meeting and no discussion:
 - a. averaging of individual's performances (statisticized)
 - b. combining of individual's performances (concocted)

This broad classification, of course, fails to consider every variant of "group." For instance, the above classification does not give appropriate consideration to the interacting non-face-to-face group which has been used to evaluate the effect of

different kinds of interaction networks, or of different kinds of information, on individuals trying to achieve some common end. Some of the network research has had as its dependent variable the quality of cohesiveness or the speed of problem solving; some of the information control studies have been concerned with group and individual satisfaction or success in completing a task. None of the research, however, has been oriented toward group vs. individual comparisons. Therefore, these important studies are not considered in the following discussion.

This review, also, omits studies concerned with group psychotherapy or group discussions designed to develop insights about individual's attitudes. Group psychotherapy in some ways, overlaps the "social climatized group" but its members do, or are expected to, interact. The individuals constituting the group are selected by an outside agent because of his belief that they as individuals will be changed by the nature of their interaction in, and with, the group. There is no group goal, but there is an individual objective for each person in it—amelioration of maladjustments and the achievement of self-understanding.

Similarly, in some forms of opinion research, the experimenter meets with an assemblage of individuals to elicit a gamut of attitudes and values about an issue. The assemblage meets for the analyst's purpose and not for the individual's, or the group's, objective, although its individuals may achieve some groupness. In group discussions for opinion research, the group provides an atmosphere which tends to facilitate individual contributions for the experimenter, but is not oriented toward either any participant's goals or a

mutual task for all of its units.

Therefore, the review is limited to researches contrasting individual with any of the six groups conceptualized, i.e.: (a) traditioned, (b) *ad hoc* (c) climatized, (d) social climatized, (e) statisticized, and (f) concocted. Although some variations may be assigned cavalierly to the nearest broad category, nevertheless, such categorization may aid in reviewing specific studies. As such, the classification may provide a clearer basis for interpretations about conclusions involving the multimeanings of *group*.

CONSIDERATIONS IN INDIVIDUAL AND GROUP COMPARISONS

History of Subjects

Performances of groups, of course, are contrasted with those of individuals. The individual, too, is a multimeaning concept. The individual may be an executive thoroughly accustomed to making policy or taking action; or the individual may be any young person selected at random from a larger population to participate in an experiment. In the same sense that the "traditioned" group survives because of its joint ability to solve problems, so, too, the functioning executive continues because of his proficiency in setting policy or in making decisions. Comparisons of groups with individuals, indeed, should give full consideration to the similarity of the experiences of the groups and of the individuals. Logically and psychologically the traditioned group should be compared with the functioning executive; the *ad hoc* group with a random individual selected from the same supply. Researches all too frequently fail to appreciate the significance of, and the need for contrasting the equivalence for, groups and individuals in the quality level of their background

or responsibility for action. Marston (44) demonstrated this in a study reported by Kelly and Thibaut (30). He showed that in the realm of legal judgments, a collection of untrained individuals may be an inferior judge of events when compared to a trained individual.

Individual Product

What performances should be involved in comparing the group with the individual? Is the group product to be compared with the average of the individual products? with the average individual? with the best individual? or with the "summed" individual? The usual procedure contrasts the average individual with the average group, although studies which contrast the *best* individual, or the "summed" individual with the best group may lead to different conclusions. Concern with the average disregards the fact that, in general, one (or more than one) individual exceeds the best group and conversely that one (or more) individual does worse than the worst group. Actually other mathematizations may be required to compare individual and group product; perhaps measurements based on probabilistic or other systems.

Motivation

In such comparisons, motivation, too, is often ignored. The view frequently cited in the literature is that meeting in a group stimulates participation and discussion, as well as interest in, and acceptance of, the experimental task. For instance, if, in a group of five members, two or more reject the experimental situation or task, the group may still emerge with some final product. By contrast, however, when an individual is not motivated to accept a situation or a

task, there is no product, or certainly not a representative one from such an individual. Some product of a group, therefore, may exceed that of an individual, primarily because of differential task acceptance.

In addition to differential motivation, there is also the possibility of differential acceptance of the responsibility in the experimental situation. The degree to which a feeling of responsibility for the decision affects the content and quality of group, and of individual, decisions is not known, but should be recognized in evaluating results. Obviously, the effect of responsibility in experimental situations cannot approximate the real situation.

Tasks

Not only do "group" and "individual" vary in setting and in motivation, but also a partial confounding, in the statistical sense, exists between task and kind of group. For instance, studies using "statisticized" groups tend to use tasks requiring estimating or judging. "Estimating" refers to estimating the number of items, the length of lines, the weight of substances, things perceptible to the senses, e.g., Bruce's (8) Ss estimated numerosity of buckshot on a card, or Schonbar's (53) Ss estimated line length.

The "climatized" group tends to be primarily used in "learning" experiments on improvement in knowledge of subject matter or the mastery of a skill, e.g., Gurnee (24) measured improvement in the mastery of a maze.

The *ad hoc* group most commonly is used in studies in "problem-solving." "Problem-solving" is to mean the thinking out of the correct answer to a problem. Shaw's comparison (55) of problem-solving by individ-

uals and by *ad hoc* groups is illustrative. Her problems were mathematical puzzles for which there is a known (or, at least, a knowable) solution. Such problems are less characteristic of life situations since few real situations occur for which the solution or the decision is known or completely knowable. The trend in research has been away from the puzzle with a Eureka solution to more realistic problems, where adequacy or goodness (not correctness) is the criterion, e.g., Maier's (41) human relations problem.

In this review, the range of problems from the puzzle to the human relations situation to the policy decision has been distinguished. While it is difficult to state wherein the processes needed to solve puzzles differs from those required to establish policy, it is felt that the nature of the potential feedback is not the same for all kinds of problems. Eureka problems can be evaluated as right or wrong, but human relations problems must be evaluated in terms of relative goodness—the range of considerations in the solution is evaluated, e.g., Maier's (40) parasol assembly problem. His problem had no correct or unique solution for adjusting the slow worker who is the bottleneck in an assembly line; rather, the several alternative plans for action must be appraised for "elegance" in terms of likely consequences.

Criteria

The truly "traditioned group" as such, has not been used in researches on the quality of group product, although some approximations have been made in studies of quantity of "productivity," e.g., Coch and French (11) measuring output of factory workers.

The multiplicity of different ex-

perimental tasks leads to a multiplicity of criteria. In "estimating," the criterion is the true order or true number; in mathematical problems or in puzzle solving, it is the right answer; in "learning," it is improvement; in "judging," and in complex problems, it is consensus of experts about the order of merit of the material or the quality of the decisions.

Side Effects

Studies of problem solving vary not only in nature of the group used, the kind of problem or task worked on, and the criterion, but they vary also in concern with the side effects of group participation. These include personal gains from the experience, commitment to decision, personality development or personal growth in empathy for the needs and feelings of others. This review is primarily concerned with studies estimating the quality of group, and of individual, products (although relevant studies of side effects will be considered). The review excludes group process as such, emphasizing only those studies contrasting the quality of the product from group interaction with the *quality* of the product by the individual.

Undoubtedly, there are many situations for which side effects are the major concern, and for which a group dynamics program has been instituted. Yet, in the military, in education, and in industry, quite frequently it is the *quality* of the group product that is the major desideratum, and frequently such side effects as commitment, morale, or feelings of participation are of less importance.

One common inadequacy of all reviewed studies is that of Ss. Most tend to use any Ss available. The absence of studies with truly functioning groups contrasts sharply with

the very large number that use college students. Among the major findings of this review of experimental work in group products has been the recognition of the relatively narrow base for the consistently broad generalizations about groups in complex situations, not only because of the narrow range of the kinds of Ss but also because of the narrow assortment of puzzles, games, riddles, and judgmental tasks. Insofar as the generalizations derive from the "statisticized," the "climatized," and the "*ad hoc*" group rather than from the functioning "traditioned" group, the generalizations in the main are founded on the behavior of college students with their less certain motivations and responsibilities rather than on the behavior of adults working under the genuine tensions and pressures of life. Thus, the generalizations will be limited and possibly not too realistic. Generalizations, psychologically, may be limited by kind of group, nature of population, the kind of task, and the basis of estimating correctness, goodness, or adequacy.

JUDGMENT

Judgment, in its long use in psychology, has often been used in research to contrast group with individual products. One type is exemplified in the work of Sherif (56), where the individual qua individual makes judgments in the presence of others to get an estimate of the effect of the group setting either upon the group's judgment or upon that of the individual. Another type contrasts the quality of judgments by the group with those by individuals to the same stimuli. In many reported contrasts of the judging of the "group" with the "individual," the "group" in the dynamic sense never

existed; rather, it was an average of several judgments made by noninteracting and separate individuals, i.e., a "statisticized" group (3).

Judgments by Statisticized Groups

The earliest use of a "statisticized" group was by Hazel Knight in 1921 (33). In her best-known experiment, college students estimated the temperature of a classroom. The judgments of the individuals ranged from 60° to 85°; the "statisticized" group judgment was 72.4°, approximating the actual room temperature of 72°. The "statisticized" group judgment was better than that of 80% of the individual judgments, even though 20% of the latter are as good as, or superior to, the "statisticized" group. In a distinctly different experiment, the Ss, in the absence of any other information, ranked 12 children for intelligence from their photographs. Each S ranked the children independently; then "statisticized" group rank order was obtained. The "group" rank order did not correlate with actual intelligence test scores any better than the individual rank orders. Finally, an *ad hoc* group of 10 Ss, met together to discuss each photograph in order to obtain ranks by an interacting deliberative *ad hoc* group. The *ad hoc* group rank order was significantly more accurate than that of either the individual or the "statisticized" group ranking. While just one *ad hoc* group was too small for generalization, Knight developed a new approach to group versus individual judgment.

In 1923, Gordon (20, 21) began publication of her series of studies extending Knight's technique of the "statisticized" group. College students ranked weights appraised against the criterion of true order.

The average of 200 correlations for that many individuals was .41. Averaging any five random individual rankings at a time, she obtained 40 "statisticized" group rankings which correlated with true order .68. Similarly "statisticized" groups of 10, 20, and 50 were computed. For four "groups" of 50, the average correlation rose to .94. Gordon, nevertheless, reports that among her 200 individual correlations, five were at least as high as .94. Her primary conclusion was that "results of the group are distinctly superior to the results of the average member and are equal to those of the best member."

The Knight technique of statisticizing group judgments was used by Smith in 1931 (59). He developed groups of 5, 10, 20, and 50 undergraduates who worked individually on the task of judging personality and behavior traits of children from written reports of their behavior. Although the correlations against the criterion (Smith's own judgment of the correct order) increased as size of group became larger, the increase was not as great as in Gordon's study. The average correlations based on 50 individuals was .37, versus .51 for the one "statisticized" group of 50. Six individuals exceeded the "statisticized" group correlation. Smith attributes the low correlations to the great number of, as well as to the ambiguity of, the traits, rather than as evidence of shortcomings in group judgment.

Judgments of weight as well as the numerosity of buckshot were made by Bruce's (8) Ss in 1935. The average of the 120 correlations for individuals with actual weights was .50; the average for two "statisticized" groups of 60 was .88. For the visually-presented buckshot, the average of the 120 correlations for individuals

was .82; for the two "statisticized" groups of 60 was .95.

Eysenck (16) used "group" techniques in 1939, when his Ss judged the beauty of 12 pictures against the criterion of the average judgment of 700 students. An experimental group of 200 was selected from the same college population. The average judgment of the entire 700 was considered the "expert." Correlations for the 200 individuals against the "expert" averaged .47; four "statisticized" groups of 50 averaged .98; and for one "statisticized" group of 200 the correlation became unity. Eysenck also reported a table of the increment in correlations as a function of number of judges utilized in "statisticized" groups.

In 1945, Klugman (31), using the Knight method, had high school students judge the number of several kinds of items in a bottle: "familiar" (jacks, marbles) and "unfamiliar" (lima beans, marrow beans). For the unfamiliar items, the one "statisticized group" of 60 was significantly closer to the true value than was the average of the individuals. On the familiar items, by contrast, there was no significant difference. Klugman concluded "when items are unfamiliar group judgment is significantly better than most individuals while on familiar items only a tendency appears."

Soldiers estimated the dates of the ending of the war with Germany and with Japan in another Klugman study (32). For the German armistice, of the 109 individuals who were tested, 27 were closer to the actual date than the "group" mean; and, for the Japanese armistice, 59 were closer. For the German armistice, he found a significant difference between the percentages of individuals with errors greater than the "group"

error as contrasted with individuals with errors less than the "group" error. This difference for the German armistice, he "interpreted to mean that the group judgment was better." For the Japanese armistice, however, there was no such significant difference, though the direction favored the individuals. In conclusion, Klugman quotes Poffenberger—"one cannot say categorically that a group opinion will or will not be better than the opinions of the individuals that comprise the group."

Not until 1932 were the obvious defects of Knight's so-called "statisticized" technique criticized. Stroop (62), after verifying Gordon's results by repeating her experiment, then adapted it by requiring just *one* individual to make 50 separate judgments, i.e., his four "statisticized groups" of 50 were four *individuals* who had each made 50 judgments of the same stimulus. When he combined 5, 10, 20, and 50 judgments of the *same* individual, he obtained correlations with the criterion nearly identical with those that Gordon reported for combining 5, 10, 20, and 50 judgments of *different* individuals. Stroop argued that Gordon's results, rather than demonstrating the social psychology of "grouping" merely illustrated an obvious statistical principle of reducing the error variance.

Farnsworth and Williams (18) in 1936 demonstrated that Knight's results were unrelated to the fact that the individual made estimations in a group setting. They repeated her experiment in every detail except that each individual estimation was made in isolation. The accuracy of their "statisticized group" results were almost identical with Knight's. In another experiment, they attempted to show that improvement

by "grouping" was not a general principle but only applied to judgments about familiar material. Using the size-weight illusion, subjects hefted two boxes and then estimated the weight of a third constructed to be lighter than either of the others although larger in bulk. The estimations were made by individuals from whose data the "statisticized group" estimations were computed. These "group" estimations did not approach the true value, leading Farnsworth and Williams to conclude that when the material is unfamiliar, distorted in a way such that all individuals are prone to make similar errors of estimation, the "statisticized" group estimation is not likely to be any closer to the true value than are individual estimations. Klugman's first study (31) indeed, was an experimental investigation of the Farnsworth-Williams generalization, but, contrary to Farnsworth-Williams, he found that when Ss are unfamiliar (as he defines "unfamiliarity") with the object to be estimated, the "group" estimation is significantly different from that of the individual.

Despite Gordon's defense in 1935 (22), the contention that "mere grouping ranks does not produce correlations," critique of her methodology was continued by Dashiell in 1935 (16), Preston in 1938 (49), and Smith (58) in 1941. The recent criticism has emphasized that, regardless of the statistical argument, experiments in which groups never meet can add little to understanding group process in social psychology. Preston (49), for instance, suggests that, notwithstanding what the Gordon results do show, they give no evidence either for psychological process or for group interaction. These studies have been cited not only because the

technique has been used widely, but also because the Gordon and the Knight studies in particular are used as evidence for the values of group process.

Judgments by Interacting Group Members

What are the results from judgments by groups with genuine interaction among its members? The earliest study is that of H. E. Burt (9) in 1920 with testimony. Individual Ss heard testimony by "stooges," some of whom were lying and some telling the truth. Each S judged which "stooges" were truthful and which were not. The individual votes were tallied and announced immediately. Ss then were constituted as a total interacting group of "jurors," who, after discussing the testimony, voted again as individuals. On the first vote, 48% were correct; after discussion, the percentages were not different. Of the 25 shifts in vote, 14 were in the right, and 11 in the wrong, direction. Burt concluded that while discussion alters judgments, it does not necessarily improve them.

Dashiell (15) reports a study by Bechtereve and Lange in 1924 in which individual judgments were made for a variety of tasks, ranging from the time interval between two sounds to the justification for a man beating a boy who had stolen from him. After the individual judgments had been made and summarized, the results were presented for discussion; after that, a second individual judgment was made. Their results seem to be consistently in favor of postdiscussion judgments. Bechtereve and Lange maintain that the group process is beneficial for all individuals, although those who have less to offer the group gain most by it. Dashiell

states that Bechterev and Lange's report is not clear about the extent of actual discussion or even if the reading of summaries itself was the discussion.

In 1932, Arthur Jenness (29) investigated the effect of discussion in *ad hoc* groups on the accuracy of individual judgments. Ss as individuals estimated the number of beans in a bottle; then they discussed the estimates in *ad hoc* groups of three, and made a group estimate; and, finally, made a second postgroup individual estimate. In two different experiments, Jenness formed the *ad hoc* groups in different ways: in one, the individuals were chosen to make for maximum disagreement in the groups; in the other, the individuals were chosen to assure maximum agreement. With *ad hoc* groups selected for maximum disagreement, group estimates were less accurate than the average of individual estimates had been, but their individual postdiscussion judgments were better in 20 of 26 instances, with a 60% average reduction in error.

In a control experiment, in a class of individuals who as individuals made two estimates without any intervening discussion, there was an average reduction in error of 4%. When the *ad hoc* groups were selected for maximum agreement, however, the group estimates were more accurate than the first individual estimates had been, but the *postdiscussion* individual judgments were not significantly different from the control. In a fourth aspect of the experiment, after the initial individual judgments had been made, the results were read to the class who were then allowed to form groups as they wished. The results parallel closely those for groups selected for maximum disagreement. Jenness (29) concludes that discus-

sion does not make group estimates more accurate, but stresses the importance of the knowledge of difference among judges in improving group judgments. He also introduced a method by which to estimate gain from group participation, i.e., the gain made by the individual subsequent to the group result, an appraisal too frequently ignored in the experimental literature despite the fact that it is usually suggested as an advantage of group process for education and for industry.

Judgments by Noninteracting Group Members

In 1937 Herbert Gurnee (23) attempted to evaluate only the effects of discussion by contrasting the judgments of individuals with those of noninteracting face-to-face groups with a sequent measure of group opinion. Individuals made their judgments on a written true-false examination. The same statements then were put to them in groups of 53, 57, 66, and 18 where each judgment was made by acclamation, with a show of hands when necessary. In each experiment, the group was better than the average individual, and approximately equal to the best individual. Gurnee computed "statisticized" groups but found that four of his five face-to-face groups were superior to their statisticized computed results. He reports a social influence upon the doubtful, in that those who were more certain of their judgments often carried the doubtful with them. His general conclusion was that although the group will be superior, the amount is unpredictable since the amount of gain depends on how well the individuals in it will do, since a task difficult for its constituent members will also be difficult for the group.

The improvement in accuracy of group judgment was demonstrated by Rosalea Schonbar (53) who reported that pairs of Ss were more accurate in estimating line length than were individual Ss since the pairs seem effective in cancelling of over- and under-estimates in interaction.

In his review, Dashiell (15) reports an experiment he conducted. He compared the written reports of two different witnesses to a staged classroom incident with those written by legal psychology students after hearing an oral version of the incident. The legal psychology students first reported as individuals their conception of the original event, and then made a subsequent report as a group. None of the seven students as individuals gave an account as complete as either of the two witnesses; most of the individual reports were intermediate between those of the two witnesses in accuracy. The group report was less complete but more accurate than either witness and all but one of the seven individuals.

Generalizations

What generalizations can be made about group and individual judgments? Generalization is more difficult than the earlier work based upon the "statisticized" groups had implied. Increase in accuracy of judgment is not obtained by the simple expedient of convening people into a group. For the results of Farnsworth and Williams (18) and Klugman (32) have shown that for some type of material a group judgment does not differ significantly from that of the average individual; and Jenness (29) as well as Gurnee (25) indicated that group superiority depends upon the quality of the judgments, and the range of judgments of individual members of the group. At best,

group judgment equals the best individual judgment but usually is somewhat inferior to the best individual. Bechtereve and Lange (7) have shown that the individuals making the poorer estimates benefit more from interaction in a group than those making the better estimates. Verbal interaction, however, does not seem essential to improvement, for Gurnee (25) as well as Bechtereve and Lange (7) obtained significant improvement without discussion. Regardless of the shortcomings of the "statisticized" group technique, the experiments with face-to-face groups also showed improved group judgment. This predicted superiority of groups is more probable when the material is unfamiliar or when there is an extensive range of opinion in the group.

LEARNING

In using learning as a basis for contrasting groups and individuals, the researches usually are less rigorous than those based upon judging or estimating. The lack of rigor comes from the ambiguity of terms, e.g., "the lecture system," "class discussion," and "study group," as well as the semantic confusion in "group" and "individual." Furthermore, the concepts of "change," "improvement," and "growth" have no reference either for "greater" or "more" or in statistical significance. Many reports are more testimonials by classroom teachers for methods they have used than experiments. As is usual in evaluating methods of instruction, variability in the quality of instructor, or of the instruction is ignored.

G. Ryan (52) made one of the first studies using learning as a basis for evaluating group and individual achievement. She divided each of four college levels into equivalent

halves for intelligence. One of each of the equivalent halves studied English and education as individuals for a six-week period with the instructor available for consultation; the other studied as a "regular" class. At the end of the first six weeks, the halves reversed roles; and for a final six weeks both halves studied together as a "regular" class. At the end of each six weeks' period, comprehensive achievement tests were given with the general result that those who had studied as a "regular" class did better. Despite such results, however, Ryan concludes that when time spent on study is equated, independent study was superior for freshmen, sophomores, juniors, and seniors and for all ability levels. This interpretation is based on the assumption that independent study took less of the instructor's time than did class instruction. Ryan seems to gloss over significant aspects of her results by implying that one goal of education is saving the teacher's time.

In 1925 Bane (5) reported results comparing the "lecture method," and "class discussion" technique. Ss were college students in education and psychology. In each of five "experiments" those taught by "class discussion" did significantly better on tests of delayed recall. On immediate recall, however, three did worse.

Using two equivalent sections of students, in 1926, Barton (6) gave each the same preliminary instruction on first-year algebra problem solving. One section was assigned new problems to solve as individuals; the other solved the new problems using class discussion. Two posttests of problem solving in algebra favored class discussion.

In 1928 Spence (61) compared the

efficiency of learning by the "lecture" system with the "class discussion" system. Two large classes of approximately 150 Ss each were compared. The first section took an initial test, studied under the "lecture" system for one semester, took a second test, then studied under the "class discussion" system during the second semester, and took a final examination. The second section followed the same schedule, except it reversed the order of "lecture" and "class discussion" study. The test results indicate superiority for students in "lecture" classes. During the first semester those who had the usual "lectures" forged ahead. During the second semester, those who previously studied under the "class discussion" method made up the lost ground. The large size of these classes should be borne in mind. These results may be valid for extremely large classes, but varying results may be obtained as class size varied.

The three following experimenters agree on the beneficial side effects of "class discussion" in comparison to the traditional "lecture" method, but disagree as to which is the superior learning system.

Thie (64), using high school English students, contrasted two equivalent halves on the basis of ability, one half working as a usual class with instruction by lecture, and the other studying in groups of five members each. Improvement was measured by difference between pre- and post-term scores on a reading test and on the writing of an original paragraph. Not only did the half that had studied in groups show greater improvement on both tests, but, in addition, these students showed greater gain in self-sufficiency as appraised by amount of voluntary work, by individual activity, and by reported enjoyment of

the course. Of the 24 students in groups, 16 registered for another term of English, in contrast with but one of the 24 of the class. Though the so-called measures of self-sufficiency are not adequately defined, Thie was one of the first to suggest that the benefits of small group techniques in the classroom may be underestimated when evaluation neglects side effects and emphasizes content achievement only.

In 1927 Zeleny (73) made much the same point in the first of his studies on the discussion-group method of teaching with college students in sociology. The experimental classes were formed into groups of seven who were given written assignments and a syllabus. The instructor gave help to the seven-member groups as needed. The control classes were taught by "traditional lecture." On terminal tests of factual knowledge and of opinion there were no significant differences between the group-discussion method and traditional lecture method. Zeleny suggests that the expected values of group discussion were not in content mastery but rather in more teacher-student cooperation, increased mutual tolerance of each other's views, and better working together with others without sacrifice in subject mastery.

In a second study of group learning, Zeleny (74) in 1940 matched two classes for age, sex, intelligence, and subject-matter proficiency. These were taught by the same instructor: one by lecture-recitation, the other in discussion groups of five students each. On gains in content knowledge, there were no differences between the two. Groups, however, were superior to those who had lecture recitation, in participation, in personality development, in social adjustment, and

in cooperation. Essentially, the results corroborate Zeleny's earlier suggestion that the advantage of group techniques for school learning is more in personality changes than in mastery of academic skills and knowledge.

In 1951 Asch (2) conducted an experiment to compare the over-all effectiveness of nondirective teaching (group participation) using the counseling methods of Rogers, Combs, Snyder, etc., to the usual lecture method. Four undergraduate sections of general psychology served as Ss. The experimental section was informed that no tests or final examination would be required. The control sections worked toward a final examination. The groups were compared for knowledge of subject matter, social attitudes, emotional adjustment, and the over-all evaluation of the course. The results indicated that the control group was superior to the experimental group in knowledge of subject matter. However, both groups were not similarly prepared to take the final examination. On the personal evaluations of the course by the students, a number suggested that nondirective teaching encourages greater amounts of outside reading, stimulates thinking about basic conceptual material, and makes for more independent decisions based on the knowledge of many individuals and not just one "authority." No differences were found between the directive and non-directive groups concerning their social attitudes as measured by the Social Distance Score. A comparison of MMPI scores indicated that the nondirective group improved to a significantly greater degree than the control group in emotional adjustment. Finally, an analysis of the Course Evaluation Forms completed

by each *S* indicated that the *Ss* felt that the experimental section was more helpful in teaching the subject matter than did the members of the control group.

In 1937 Gurnee (24) reported the first of two "learning" experimental studies which, though more rigorous in control of basic variables, used somewhat less realistic problems. Half of his *Ss* worked as individuals, while the other half worked in groups averaging 10 members. The task was to learn a maze. Individuals were to concentrate on eliminating errors without concern for time; groups voted each step in the maze by acclamation. Groups and individuals had six trials. Groups did significantly better than individuals by having fewer errors and completing the first perfect trial sooner. Gurnee, then, tested all *Ss* as individuals on a seventh trial. He did not find any significant difference as a result of the two different kinds of experience. His results may be subsumed under Jenness' generalization that when groups are in agreement, group members will not improve as a function of group experience.

The next year Gurnee (25) reported a similar study with quite different results. In addition to the same maze, he added the learning of the arbitrarily correct number of 20 pairs of two-place numbers each in the course of six oral trials. Individuals were contrasted with groups, then a seventh written trial was given to all as individuals. On the seventh trial, those who had worked in groups made significantly fewer errors than those who had worked the first six trials as individuals.

Moore and Anderson (45) contrasted the learning of *ad hoc* three-man groups with that of individuals in applying some of the laws of the

calculus of propositions in order to solve 10 different symbolic logic problems. The results are based on six individuals and six matched groups. In general, none of the differences is significant; i.e., the number of steps taken, of errors made and of time to solution did not differ statistically between individuals and groups. There was a greater tendency, however, for individuals to repeat steps, suggesting that the members of the group remember steps taken. It is not surprising that estimates of variance for groups and for individuals usually are not significantly different, but the direction is always for greater variance among individuals. Nevertheless, in the use of symbolic logic Moore and Anderson have introduced a novel learning task for use by psychologists.

Contrasting the results by groups and by individuals in "learning" suggests quite amorphous generalizations. Spence indicated that the lecture system is superior to the class discussion system for large classes. Ryan's results agreed. Asch's results, in a narrow sense, must be similarly interpreted. Thie, on the other hand, found that under his experimental conditions "class discussion" produced significantly better learning than the "lecture" method. Ryan found class discussion superior for certain types of learning, and inferior for others. Zeleny and Gurnee found no significant differences between the two forms of learning. These amorphous results suggest several explanations, the most likely of which is that these experiments were conducted under such varying conditions that seemingly diametrically opposed results are understandable. For example, the size of groups can be expected to have a profound effect on results. It is known that as group size in-

creases, individual involvement decreases, and inhibition increases. Large discussion groups, therefore, might be expected to produce less learning than smaller groups. Other factors such as announced goals, subject matter, methods of measuring improvement, etc., can be expected to have profound effects on the results. This indicates a serious need for additional experimentation which can control important conditions.

SOCIAL FACILITATION

Social facilitation refers to the effects on an individual of working at a task in the presence of other individuals but independently of them, i.e., not interacting with them, although they may be face-to-face in an audience or classroom. In social facilitation experiments, there is no interaction or cooperation, and, often, no expressed feeling of rivalry, although competition may affect the results. Allport (1) used graduate psychology students, who, in a first period, took a free association test alone, then, in a second period, took it individually in "groups" of three to five persons. Fourteen out of 15 individuals produced more words while working in the social setting than when working alone, although the differences were not significant statistically. Allport found that the effect of the so-called co-working "group" on individual productivity was to increase quantity but decrease quality of the associated words. He concluded that some tasks may be better done alone than in groups.

His basic technique was used through the years by Sims (57), Sengupta and Sinha (54), and others with almost no variation in results. The assigned tasks were always repetitive and meaningless, e.g., letter cancellation in running text, etc.

In Sengupta and Sinha's study, for example, in Ss that worked at a task for nine days, output did not vary much after the third day. Upon changing the work situation into a social setting, output rose significantly until restabilized at a second but higher level. Mukerji (46) found that with children doing letter cancellation and letter-naming, almost 90% of the individuals had superior outputs in the social setting; but that oscillation in production was greater when performing in groups.

In 1952 Wapner and Alper (69) described a restraining force as a result of an audience. One hundred twenty Ss were tested in three varying situations. All were asked to select one of two words which best fit a given phrase. In the first situation only the S and the experimenter were present. In the second situation, the S and the experimenter were present, but the S was informed that an "unseen" audience was listening to and watching his performance. In the third situation, the S and the experimenter were present with a seen audience. Either task-oriented or ego-oriented instructions were given the Ss. In the task-oriented instructions the Ss were informed that the material rather than the S was being studied. In the ego-oriented instruction the Ss were informed that the task was a form of personality test and that they, rather than the task, were being evaluated. The results indicate that the time to make a choice was longest in the presence of an "unseen" audience under both forms of instructions; next longest in the presence of a "seen" audience; and shortest when there was no audience. The significant differential effects of the audience variable occurred for the first half of the experimental sessions only. Items with

personality references yielded longer times than neutral items. Contrary to the expectations of the experimenters there were indications that time to make a choice is longer for task-oriented than for ego-oriented instructions.

An extensive inquiry into the effect of co-workers on productivity is reported by Dashiell (14). He investigated the conditions of working: (a) alone; (b) together but noncompetitively; (c) together and competitively; and (d) alone, but under observation. Speed increased for each of the three tasks (multiplication, serial association, and a mixed relations test, particularly between conditions (a) and (d). Accuracy, on the other hand, was much more evenly distributed among the four conditions: with the only clear difference in the "observation" condition for which the work was least accurate although the greatest amount was produced.

Kelly and Thibaut (30) reported a study by Wyatt, Frost, and Stock (72) in 1934 which indicated that in real life situations involving work of a highly repetitive nature, social facilitation effects have been found consisting of closely similar production curves for employees working together. The authors found that workers' rates of output varied with the output of others in the work group. This relationship was particularly close for pairs of workers seated opposite each other, and was somewhat more marked the more visible and the more measurable the output. When individual workers were subsequently isolated, the correspondence between their work and that of the others disappeared.

Hilgard, Sait, and Magaret (27) indicated that not only can actual production be affected by social facil-

itation, but also level of aspiration. The Ss worked in groups of three to six members. They individually worked on successive subtraction of three place numbers. The material was graded in difficulty in order to produce experimental differences in success. After the first experimental session, when all Ss' scores were known to each other, they were asked to estimate their future performance. Those Ss ranking superior in relation to their social group tended to estimate their future performance too low, while those Ss making inferior scores tended to estimate their future performances too high. Though the critical ratios were low and caution was recommended in interpretations, the trends within the groups were clear. The authors speculated that the desire for social conformity might well produce this regression of predicted scores toward the mean.

To a degree, the influence of co-working members, indeed, may be stronger in a group interacting for a common objective. The feeling of ego-involvement in the group's product may be a significant factor as in problem solving for a group result.

PROBLEM SOLVING

In problem solving, few experimental studies contrast the *quality* of solutions by groups and by individuals. Most results seem to be by-products of investigations of the problem-solving process. Nevertheless, these few studies are those most frequently cited as evidence of group superiority, e.g., Watson's (70) comparison of groups and of individuals in problem solving. He used *ad hoc* groups of college students given the task of making as many shorter words from the letters of a larger word as possible within a time limit. For the first trial, Ss worked as individuals;

for the second and third trials they worked in 20 *ad hoc* groups of from 3 to 10 members; and for a final fourth trial again as individuals. The group product, i.e., the number of different words, was significantly larger than that made by the best individual and thus, obviously, larger than that of the average individual. When Watson (70) formed what may be called the "concocted group" or "summated individuals," i.e., added together all the different words in the first trial made by the individuals comprising the groups in Trials 2 and 3, he found the average "concocted group" product significantly larger than the average *ad hoc* group product. Even though the average product of the *ad hoc* groups significantly exceeded the product of the average individual or of the best individual, nevertheless it was significantly inferior to the full resources of all of its individual members. Group interaction may inhibit the fullest potential contribution by its members. Indeed, the superiority of the "concocted" group over the interacting *ad hoc* group suggested such an inhibition.

In a subsequent study Watson (71) evaluated group and individual superiority on nine different tasks: finding antonyms, solving a cipher, drawing conclusions from stated facts, completing sentences, listing steps in problem solving, composing limericks, comprehension of reading, and an intelligence problem. There were three equivalent forms of each task; Ss first did one form as individuals; second, another in *ad hoc* groups; and third, the remaining form as individuals. On all nine tasks, the average achievement of groups was superior to that of individuals; the differences, however, ranged from small and insignificant for reading com-

prehension to large and significant for completing sentences. In speed, on the average, groups were superior to individuals. For the nine tasks, on the average, about a third of the individuals were superior to their group in score and in speed. Such superiority, however, was a function of the task: for instance, on antonyms, 11% of the individuals made scores superior to their group in contrast to 50% of the individuals who did better than their group on the intelligence problem. The order of group superiority is as listed above.

In 1932 Shaw (55) compared groups with individuals in the rational solutions of complex problems. A class in social psychology was divided into halves. In the first period, half the class worked in five *ad hoc* groups while the others worked as individuals; in a subsequent period the roles of the two halves were reversed. In the first period, the task was the solution of three very similar classical "mathematical recreations" puzzles, e.g., the three beautiful wives and their jealous husbands who had to cross a river by rowboat carrying three persons at most, under the constraints that no wife and all husbands can row and that no husband would allow his wife in the presence of another man unless he was also present.

In the second period, the problems were quite different: (a) rearranging words to form the last sentence of a prose passage; (b) rearranging words to form the last three and a half lines of a sonnet; and (c) to find the most economical routes for two school buses to bring children to a common school under the constraint of maximum bus capacity and of a specified number of pick-up stations. For the puzzles, obviously, there is just one right answer; for the word rearrange-

ments, however, the correct answer is arbitrarily the original word order, and for the school bus problem, it is the one that gives minimum mileage. Ss are more likely to be able to verify the solution for the three mathematical puzzles which were given in the first period; but for the second period problem, they have no way of verifying their solutions because the correct answer was arbitrary. For instance, word rearrangements can be completely appropriate in meaning despite deviation from the original word order. Puzzles having unique solutions may be termed "Eureka" since Ss can, and do, get confirmation for correct solution.

For the first period, on the so-called Eureka problems, three of the 21 individuals and three of the five groups solved the first problem; no individual and three groups solved the second problem; and two individuals and two groups solved the third problem. No individual solved more than one problem, but just three groups made the eight group solutions. Two groups and 16 individuals never solved any of the three puzzles. For the second period problems, three of 17 individuals and four of the five groups solved the first problem completely; a fifth group and seven other individuals made just one error. No individual and no group solved the other two problems. Group superiority rests only on the eight solutions by groups in contrast with the five by individuals. In general, interpreters of the Shaw experiment have disregarded not only the similarity among the three problems but also the fact that the solutions were based on the sum over-all problems rather than on the number of *identical* solutions by individuals and by groups. For instance, when only the solutions by individuals and by groups for the first

problem are compared, there is no statistically significant difference. Shaw neither discussed the fact that two of the groups never solve any of the three puzzles, nor the relative efficiency of three solutions among 21 individuals versus three solutions for five groups of four members each, i.e., 20 individuals altogether.

Shaw advanced methodology by her more rigorous procedures for studying problem-solving of individuals and of groups; however, the interpretations implicit in her conclusions do not conform to the constraints placed either by the kind of problems, or the type of Ss, or the possibility of transfer of training. The fact that two groups never solve, and that three groups get eight solutions, suggests two hypotheses for research: (a) that transfer of training is more likely in groups and (b) that group solution is possible only if at least one individual as an individual could have solved the problem.

Shaw accounted for group superiority on the basis of observations that groups rejected incorrect solutions and checked against errors. Since her results differed with the different problems, her interpretations might have been that for problems with just one unique answer, groups were superior; but for problems with a wide range of answers, there is no genuine difference. Lorge and Solomon (38) re-examined the data for the Eureka problems in 1955 and suggested other explanations for group superiority. Their work is reported later in the section on mathematical models.

The question of the relative efficiency of three solutions among 21 individuals versus three solutions for five groups of four members each, i.e., 20 individuals, was investigated by Marquart (42) in 1956. She essen-

tially replicated Shaw's experiments with similar results. However, Marquart noted that Shaw's conclusions about group superiority hinge on comparing percentages of possible successes obtained by individuals to percentages of possible successes obtained by groups. A fairer comparison, she proposed, involved treating individual successes on a group basis, e.g., if, when working individually, one of the three individuals who later make up a group of three get the correct answer, individuals are credited with one success in one trial, rather than one in three. If, on the other hand, no correct solution is forthcoming from any of the three individuals, then one failure is attributed instead of three. On this basis, the individuals turned out to be slightly superior in both Marquart's results and in Shaw's.

Shaw, however, did not consider her conclusions limited by problem type. In 1938 Thorndike (65) investigated the hypothesis that as the range of responses increased, the superiority of the group over individuals increased. Thorndike used two versions of each of four problems; one with a "limited" number of responses and the other with an "unlimited" number. For instance, a multiple-choice item with four options was paralleled by an open-ended version; or similarly, completing a crossword puzzle was paralleled by requiring the construction of a crossword puzzle. The other tasks were limerick completion (either one line or three to be supplied); and a vocabulary test of synonyms, five choices or recall. Ss were college students, who worked four two-hour sessions, each a week apart, in two sessions as individuals and in two as groups. For all four tasks, differences were in the direction of the hypothe-

sis, with three of them significant.

Thorndike's tasks, in a sense, contrast recognition versus recall in groups and in individuals. The recognition item form favors groups. This indicates, as Shaw had suggested, that group superiority results more from members pooling information by rejecting incorrect options than by contributing options for consideration. Thorndike's problems differed so much from those of Shaw as to suggest that generalizations about problem-solving and about group superiority seem to depend upon the nature of the tasks.

Husband (28) in 1940 attempted the study of a group in contrast with an individual as measured by required man-hours to arrive at a solution and the quality of the product. He used three tasks: deciphering a code, solving a jigsaw puzzle, and solving arithmetic problems. Ss were students in psychology, 40 working alone, 80 in pairs. Some pairs were friends; some strangers. Pairs were superior on the first two tasks, but on the third (arithmetic problems) there was no significant difference. Husband suggests that on the arithmetic task one member of the pair tends to take the lead and do all the work. In all comparisons, pairs of strangers did better than pairs of friends.

Husband's results emphasized the conclusions from some of the earlier studies about originality and routine performance. His pairs did better on problems requiring some originality or insight than on the more routine arithmetic problem; this confirmed Thorndike's hypothesis that the superiority of the group product over the individual product is greater in problems with unlimited solutions than in those with limited alternatives and confirmed Watson's and Shaw's findings that the group han-

dled complex problems adequately. Regarding efficiency, however, he indicates that the time saved in pairs was never more than a third—not the half needed to equate time for pairs and for individuals, although Husband failed to consider the better quality for the time used by pairs.

After a long interval following Thorndike's work, Taylor and Faust (63) compared individuals and *ad hoc* groups in solving the identity of a topic in the game of Twenty Questions. Elementary Psychology students worked for four days at the rate of four problems a day, either as individuals, or in pairs, or in groups of four. On the fifth day, all Ss worked alone. Although time was recorded, the prime criterion was the number of questions necessary to reach a solution. In pairs and in groups, discussion was allowed, with the motivation that they were competing against other groups but not against each other. There were significant differences between the scores of individuals and those of pairs and of groups in questions, time, and failures. Except for failures, there were no significant differences between the pairs and groups. Of course, in efficiency, i.e., the number of man-hours to reach a solution, the group is inferior to the individual, with four-man groups less efficient than pairs. The gain from training acquired the first four days by each individual as measured on the fifth day did not seem to be different whether the first four days' training came by practicing as individuals, or in pairs, or in groups.

Taylor and Faust's Twenty Questions approximated the Eureka, but also it was summative in that each member's contributions could add to the group result. Their data tend to corroborate Shaw and Watson; but

they contradicted some studies of "learning" insofar as there was no transfer to individual achievement as a consequent of previous differential group or individual experience.

Research contrasting group and individual performance in "learning" suffered from a lack of experimental controls; research with problem solving suffered from a lack of reality, etc.; problems or tasks are far removed from the genuine and the real. The problems, in general, have been puzzles, riddles, or information-test questions. Results from such tasks were not sufficiently conclusive to allow an unambiguous generalization about the superiority of groups over individuals with more realistic problems.

MEMORY

Little work in the area of group and individual memory has been completed. In 1952 Perlmutter and de Montmollin (48) experimented with group vs. individual learning of nonsense syllables. Twenty groups of three persons each were required to learn equivalent lists of nonsense words. One list was learned by each individual while working alone, but in the presence of the other two. A second list was learned as a cooperative three-person project. Half of the Ss worked first as individuals and half first as members of interacting groups. On all trials the average group recalled more words correctly than did the average individual. The group recall tended to be equal to or better than the best individual score, and those who worked first as members of interacting groups tended to do better as individuals than those who worked first as individuals. The converse was not found to be true. In agreement with Shaw (55), Perlmutter and de Montmollin noted proc-

esses of rejection and evaluation operating within the groups, and the results seemed to indicate that groups adopted fewer invented words and fewer words represented modifications of those in the lists.

In 1953, Perlmutter (47) tested group vs. individual memory of "meaningful" material. A story entitled "War of the Ghosts" was read to 8 two-man groups, 8 three-man groups, 3 four-man groups, and 10 individuals. A comparison of recall was made after 15 minutes and after 24 hours. No statistically significant differences were found, although the results favored the groups. The standard deviations of individual's scores were nearly twice those of the three-man groups, indicating the possible existence of a group pressure toward conformity. Individuals required less time than both two- or three-man groups at a statistically significant level in both sessions. Perlmutter concluded from this experiment that, on the one hand, hardly any evidence was found to support the extreme position that the content of group memory product is unique and not related to the content of individual member recalls. Very little correct information was found in group recall that was not in any member's recall. Conversely, some correct content was found in all or some of the individual's recall that was not found in the group memory product. He concluded that while it was interesting to attempt a derivation of group product from individual products, in some respects group product can be treated in its own right, and that some principles of product change can be formulated without measurement of individual member memory.

The single generalization derived from these studies is that in conform-

ity with other studies reported in this review, evidence of the existence of both a depressing and an accelerating effect from group participation is evident. These experiments do not aid in identifying or quantifying these effects.

SIZE OF GROUP

A section on group size is included because of its profound effect upon group productivity. "Group" in contrast to "individual" is affected by a number of important variables, size being one of the most important. This is true to such a great extent, that the term "group" can refer to materially different entities. It is important that knowledge of the variability of a group product, as conditions vary, be utilized when comparing "group" and "individual" products.

In 1927, South (60) conducted an experiment with 1,312 Ss divided into groups of three and of six. Four types of tasks were assigned ranging from the "concrete" to the more "abstract." The tasks were: judging emotion from a series of photographs portraying emotion (abstract); answering multiple choice questions (concrete); solving bridge problems (concrete); and judging English compositions (abstract). The results were obtained on the accuracy of the performance, and on the time required to complete the experiment. The results indicated that the size of the group affects its efficiency. In each of the four types of material there was a difference between the performance of groups of three and groups of six, depending somewhat on the type of material or the kind of problem given the group. The small groups were more efficient with "abstract" problems, while the larger groups did better with the "concrete"

problems. South concluded that in the case of the abstract materials, the members had their own opinions after the first glance and the committee's task was largely that of compromising and overriding opinion. In the case of the smaller group there were fewer opinions and hence less to do. In the case of this particular type of problem, the small group was faster with no loss in accuracy.

Kelly and Thibaut (30) reported a study by Bales et al. (4) in which individual Ss were first ordered according to initiation rank, which is the degree to which they initiated responses in a group situation. Theoretical curves based on a harmonic distribution were then fitted to the obtained percentages of total acts contributed by members at each ranked position. For groups of size three and four the empirical curves were found to be flatter than the theoretical curves, but for groups of size five through eight the empirical curves were steeper than the theoretical curves. Thus it appears that the proportion of very infrequent contributors to the group interaction increases as the size increases. In the larger groups the discrepancy between obtained and expected frequencies was attributed to the large volume of participation by the highest initiator. This study suggested that as size increases from three to seven there is a sharp rise in the proportion of members who contribute less than would be expected if each member shared equally in the interaction. Beyond the size of seven, the proportion shows no consistent increase or decline.

In 1951, Gibb (19) experimented with the effects of group size upon idea production in a group problem solving situation. The Ss were 1,152 college students composed into groups

of 1, 2, 3, 6, 12, 24, 48, and 96. The groups were asked to produce as many solutions as possible to a series of problems permitting multiple solutions. Each group session lasted 30 minutes. The results indicated that the number of ideas produced increased in a negatively accelerating function of size of group in each of the two conditions. Valid criticism of this experiment is that the time limit of 30 minutes was not sufficient to permit an exhaustion of the potential contributions of all of the members of the larger sized groups. Furthermore, the problems may not lend themselves to more than a limited number of solutions. However, of extreme importance is that Gibb reported that with increasing size a steadily increasing proportion of the groups' members reported a feeling of threat or inhibition of their impulses to participate. This, in addition to the statistical results, supports the hypothesis of a restraining force resulting from increased size of groups.

Carter et al. reported a study (10) comparing individual participation in groups of varying sizes. They concluded that in groups of four, individuals have sufficient space in which to behave, and thus the basic abilities of each individual can be expressed, but in the larger groups only the more forceful individuals were able to express their abilities and ideas, since the amount of freedom in the situation was not sufficient to accommodate all the group members.

From this limited number of studies certain tentative generalizations can be made. As indicated by South, greater production on "abstract" problems can be expected from smaller groups than from larger ones, and greater production on "concrete" problems from larger groups

than from smaller ones. Bales et al. (4), Gibb (19), and Carter et al. (10) indicated the possibility that groups of increasing size will increase production at a negatively accelerating rate for problems of certain types. When comparing production of groups of varying size and individuals, these generalizations should be kept in mind. Considerable additional research is needed to confirm or refute these expectations.

PROBLEM SOLVING IN MORE REALISTIC SETTINGS

Studies using genuine and significant situations are less numerous than those involving judging, learning, etc., partly because concern with the more genuine human relations problems has emerged quite recently and partly because of the practical difficulties in working experimentally with problems involving decisions. These decisions require the individual or the group to weight alternatives for relative adequacy, followed by the selection of one or some combination of several as the most feasible solution rather than determination of the correct answer. Thus, the criterion for appraising decisions in these experimental studies should differ from agreement with the *one true order or the one correct answer*; rather, the evaluation of decision, ultimately, should be based on some system of credits for coverage and adequacy.

Timmons (67) used as criterion the experts' rank-order of five possible options to the genuine problem, "What type of parole system should Ohio adopt?" His research was oriented primarily to estimate the effect of discussion on the individual's ranking of the five options. The Ss were high school students in Ohio. Classes were divided so that some Ss worked

as individuals throughout the experiment and others worked in specially constituted groups during part of the experiment. The controls (as individuals throughout) 1: on Day 1, ranked the five different options and took an attitude scale toward parole; 2: on Day 2, read a pamphlet containing authoritative information about parole, then again took the attitude scale and ranked the options; 3: on Day 3, reread the information pamphlet under motivation of competing with groups discussing the problem, and then again took the attitude scale and ranked the options; and 4: after an interval of a month, were measured for attitude and for ranking of options.

The experimental section was treated identically for Steps 1, 2, and 4. The essential difference was in Step 3 in which six different kinds of groups were formed, based on the performance on the first day. Each group was supplied with a copy of the informational pamphlet, discussed the problem in leaderless groups and formulated a ranking of the options as a group. When that had been completed, each member of each group took the attitude scale and ranked the options as individuals.

Timmons' measure was the individual's ranking of options, so much so that Timmons considered ranking by groups only incidentally and tangentially. In terms of the individual's agreement with expert ranking, the informational pamphlet produced a tremendous shift toward experts' rank (Day 2 minus Day 1). Individual study and group discussion resulted in further movement toward expert ranking (Day 3 minus Day 2). The individuals who participated in group discussion were closer to the experts than the indi-

viduals who restudied the pamphlet. These changes were maintained, in general, a month later. For attitudes, gains only followed the reading of the pamphlet on Day 2 and at no other time, and showed no difference at any time between those who discussed and those who restudied the pamphlet.

In this major aspect, Timmons demonstrated a significant transfer from group discussion to subsequent individual rankings. Unfortunately, Timmons considered the group's ranking a very minor aspect of his research. He reported that after discussion the groups' average agreement score was 2.93, which was not significantly different from the immediately subsequent individual (from those groups) average agreement score of 3.31. The 3.31, however, was significantly better than the 6.70 of the individuals who had restudied the pamphlet.

Although Timmons formed six different kinds of groups based on the amount of their agreement with experts' ranks initially, he failed to report the ranking of the various groups. Methodologically, the six groups were made up as: I. 4 Ss with good scores; II. 4 Ss with intermediate scores; III. 4 Ss with poor scores; IV. 2 Ss with good scores, 2 Ss with poor scores; V. 2 Ss with good scores, 2 Ss with intermediate scores; and VI. 2 Ss with intermediate scores, 2 Ss with poor scores.

He reported, however, in terms of individual change that the gains were largest for the poor, smaller for the intermediate; and least for the good, student. The *good* made greater gains after discussing with other *good* than after discussing with the *poor* or the *intermediate*. The *good* did not get worse after discussing with the so-called *poor*. The *poor* gained as much

from discussion with the *good* as from the *intermediate*, but always significantly more than from discussion with the *poor*. From the viewpoint of learning by individuals, all individuals seem to benefit from discussion even when the discussants were relatively less adequate.

In 1941, Robinson (51) investigated the effects of group discussion upon attitudes toward two social problems: capital punishment, and American policy to keep out of war. He contrasted college sophomores in 43 *ad hoc* groups of from 8 to 20 members as experimental samples, and 225 college sophomores as individuals. The experimental sample (a) studied group discussion theory for one month and had weekly practice discussions, (b) studied material on the problems, then, (c) took Thurstone Attitude Scales relevant to both problems, (d) had a two-hour discussion on each of the questions, and, finally, (e) took the attitude scales a second time. The control sample had no group discussion, but were given successively the two forms of the Thurstone Scales on each problem. In one variation, an information test was added before and after discussion; in other variations, discussion theory and the study of the informational material were omitted.

Although significant changes in mean scores were made only in attitudes about how to keep out of war, Robinson noted that a consideration of the magnitude of the attitude shifts by individuals revealed significant changes on both problems in all groups. When the informational test was used, the individuals showed gains after discussion. However, without comparable data for the control, this gain cannot be referenced to individual versus group superiority. In a third experiment, Robinson,

comparing change in attitude after reading informational material with that after 30-minute group discussion found that the magnitude of the shifts by individuals after reading not only exceeded those after 30-minute discussion, but also those after the two-hour discussions in the earlier experiments. This lack of shift after discussion, contrary to Timmons' findings, may be attributable to the inadequacy of experimental designs, or to a genuine difference between the sequelae of reading or of discussion.

Robert Thorndike (66) hypothesized that much of the superiority of groups over individuals was attributable to the elimination of (a) individual chance errors and (b) those errors differing from individual to individual and from time to time. The second hypothesis had been confirmed partly in the Gordon experiments. Thorndike attempted to isolate that part of group superiority that was due to averaging or summing individual contributions, from that part due to the elimination of each individual's chance errors. He used 1,200 college students formed into 220 *ad hoc* groups of from four to six members. They worked on 30 problems, e.g., selecting the better of two poems, the more socially significant of two headlines, etc. The design required choice by an individual together with a measure of confidence; then, after discussion, a group choice. Both the individual and the subsequent group choice were completed for each problem before the Ss proceeded to the next problem. There were significant differences between the mean scores for all individuals before discussion, for "concocted" groups before discussion, and for *ad hoc* groups after discussion.

The analysis of the group product

revealed that part, but not all, of the difference between group and individual results can be explained as a consequence of the pooling of the individual products. As Gurnee found in group judgments, the group product was more than an expression of the majority of the members comprising the group, and Thorndike found this "more" attributable to the discussion among the group's members.

The data also were analyzed for the consequences of grouping, i.e., the effects when the majority were correct before discussing in the group, as opposed to when the majority were incorrect. When at least 70% of the individuals were correct before discussion, there was a gain after group discussion of 11%. When less than 50% of the individuals were correct before discussion there was a loss after group discussion of 7%. This result indicated the necessity of qualifying Jenness' earlier hypothesis that disagreement among members is more conducive for group improvement than is agreement. His hypothesis may be correct with judgments, where awareness that others do differ results in restudy, but may not apply in problem solving or decision making when disagreement involves a majority with an erroneous view. Indeed, Farnsworth and Williams (18) made the same point in their demonstration that when all group members were likely to be in error, there was no reason to expect the group product to be *better* than that of individuals.

Timmons (68) concluded that after allowance for the averaging of individual contributions is made, a significant superiority for all groups still remains; similarly, after allowing for the effect of majority influence, an insignificant amount of superiority is reported. When allowances both for

averaging effects and for majority influence are made, there is "a large, but not significant" difference favoring the group. Timmons suggested that considering the rigor of his methods, "it seems probable that the differences are even more significant than they seem to be." He suggested four factors possibly inherent in discussion that may account for unexplained differences: the group (*a*) has more suggestions leading toward the solution (cf. Shaw), (*b*) has a wider range of interpretations of the facts of the problem, (*c*) has a wider range of criticism and suggestion (cf. Shaw), and finally (*d*) has more information (cf. Lorge and Solomon).

In 1955 Lorge et al. (39) experimented with the difference in the quality of solution to a practical problem which was presented in four settings differing in degree of remoteness from reality. The problem was presented either as a verbal description, a photographic representation, a miniature scale model, but not allowing manipulation of parts and materials; or a miniature scale model allowing manipulation of parts and materials. The problem consisted of finding a way to get a squad of soldiers across a specified segment of mined road as quickly and secretly as possible using a limited number of available props. The problem was adaptable to many solutions of varying quality, and had the characteristics of a genuine field situation. Ten teams of five AFROTC students and 10 individuals worked on the problem. Any individual or group had the right to ask as many questions of the experimenter as he so desired, and the experimenter made an effort to answer all questions as long as they did not directly divulge a method of solution. The results indicated no significant differences

among the solutions at the four levels of remoteness from reality. However, at all levels the solutions of the groups were markedly superior to the solutions of the individuals. It was concluded that the differences may be due in large part to the amount and kind of information the Ss had at their command. It was noted that the number of questions asked of the examiner increased with the remoteness of the model from reality. This worked toward equalization of the information available to all groups. Groups asked more questions than the individuals at all levels of remoteness, which meant that the groups had more information to work with than did the individuals, and may account, in part, for their superiority.

Lorge et al. (36) completed a study in 1953 comparing the quality of group and individual solutions of human relations problems before and after class instruction in staff procedures. At the beginning of the course, one half of the Ss spent one period in problem solving as individuals, while the other half worked as groups. In the very next period, those who had worked as individuals were formed into groups, while the earlier groups were dissolved and their members worked individually. The design was replicated six months later, at the termination of the course. The results indicated that as a result of this particular form of instruction, the quality of decisions prepared by *ad hoc* staffs after training is significantly superior to that of those prepared by *ad hoc* staffs in the opening week of class. By contrast, and of interest, the decisions written by individuals after instruction are not significantly different from those they had prepared as individuals in the opening days of the

course. This indicated the possibility of individuals being able to improve in their performance of a given task as members of a group, without showing improvement in their performance of the same task as individuals.

These results emphasize the fact that the group is not necessarily superior to the individual in human relations decisions. The quality of individual decisions before instruction is significantly superior to that of groups. This difference in favor of individuals may indicate merely the relative ineffectiveness of *ad hoc* groups to solve the problem in the given time. In some of the appraisals it was found that groups, before instruction, lost upwards of 80% of the ideas that their constituent members as individuals had for the solution of the problems. Many of these lost ideas were important. At the beginning of instruction, more than 75% of the individual decisions were superior in quality to the best of group decisions. Since the decisions of the individuals at the end of instruction do not differ in quality from those at the beginning, the presumption was that there was a gain in group interaction but not in problem-solving skills whether among individuals or in groups.

The data also indicated that the probability that any individual's idea will be expressed in the group decision is a function of the *commonality* of the idea, i.e., of the number of individuals who had the same idea *prior* to the group meeting. Of all the ideas that were held in common by two or more group members prior to the group meeting, half appeared in the group decision. Only 10% of the unique ideas, those possessed by only one person prior to the group meeting, ultimately appeared in the group decision. Similarly, only one

third of the ideas evolved in group decisions were original, i.e., ideas which none of the group members had mentioned in their earlier individual decisions, whereas two thirds of the ideas had already been so expressed. These data suggested that group process does not generate original ideas but relies heavily upon ideas formulated prior to the group meeting.

A further aspect of the group's involvement with decisions was in the changes in food habits studied under the sponsorship of Lewin (35). The basic comparison was the carrying into actuality of decisions made either in groups or as individuals. As such, the Lewin studies concern a side effect of the group process without any reference to the quality of the decision.

One study was conducted with six groups made up of Red Cross volunteers in a home nursing course. The objective of the course was to increase actual use of the unpopular "variety meats," e.g., beef hearts, kidneys, and sweetbreads. Lecture was contrasted with group discussion in each of three groups to induce change, with the same nutritionist offering the same recipes. At the end of the discussion, the women, by a show of hands, indicated whether they intended to try the new foods. On follow-up, 3% of the lecture and 32% of the group-discussion volunteers used some of the "variety meats." Lewin stated, however, that only subjects in discussion groups were told of the planned follow-up.

In a study by Radke and Klisurich (50) six neighborhood groups composed of from six to nine housewives were organized with the objective to increase home consumption of whole and evaporated milk. The contrasted procedures were lecture and

group discussion. After two and after four weeks, follow-ups indicated that those who had the discussions showed significantly greater change in the desired direction. As in the Lewin study, discussion groups were informed of the two-week follow-up but the lecture groups were not. Lewin, however, stated that neither sample knew about the four-week follow-up. The differences, on the four-week check, may be a consequent of successes with the new foods tried.

Klisurich and Radke (50) tried to have new mothers increase the amount of orange juice and cod liver oil fed their babies. The "individual" condition involved each mother individually discussing with the hospital nutritionist for about 25 minutes the feeding of her new baby after which she was given printed instructions on feeding. Both oral and written material stressed the importance of using orange juice and cod liver oil. The "group" involved other new mothers who were formed into *ad hoc* groups of six members for instruction and discussion of feeding. The time for a group of six was equivalent to that given any one individual, i.e., 25 minutes. Follow-up was made after two and after four weeks. The results show that significantly more who had made decisions in groups behaved in the desired fashion.

Lewin suggested that the first two experiments may be interpreted as the consequences of (a) greater involvement in the group situation as compared with the more passive audience role in the lecture, or (b) greater interest in the group discussion or (c) that only those in the discussion groups knew of the anticipated follow-up. The results of the third experiment were all the more striking because the individuals re-

ceived special attention, much more than any group members; and, further, because the group members were farm mothers who were unacquainted with each other before and who had no subsequent contact after leaving the hospital. Nevertheless, a 25-minute discussion among six such strangers produced much greater change than did a 25-minute consultation with an individual. Lewin considers the third experiment as indicative either of greater individual involvement in the group decision, even under the described conditions, or that decision in groups tends toward action.

Levine and Butler (34) replicated the Lewin experiment, attempting better controls, i.e., avoiding differential expectations. Their Ss were factory supervisors who regularly gave ratings which determined the base pay of workers in their departments. It had been shown previously that the supervisors tended to rate the job rather than the man, i.e., workers for the more highly-skilled jobs were consistently rated higher than those for the less skilled jobs. The experimenters educated the supervisors to rate man performance, not job level. The 29 supervisors were randomly divided into three groups. One group had no training, one a 90-minute discussion on improving rating, and the third a lecture followed by a question and answer period totaling 90 minutes.

The control group did not change, rating the men on the more skilled jobs higher than the men on the less skilled jobs. "The lecture method had practically no influence upon the discrepancies in rating.... Performance ratings were affected significantly only after the raters had had a group discussion and had reached a group decision." The data,

however, are not sufficiently rigorous to supply definite evidence for the superiority of the group technique.

In 1950, Maier (41) compared groups and individuals in decisions with a realistic problem. The task was to plan the action to solve the problem of what to do with the slowest man of a "parasol" or circular assembly line. The slowest men held up the whole assembly line. The problem was presented in two ways: (a) just a statement of the problem or (b) added to the statement of the problem was a description of the roles of the eight different members of the "parasol" assembly.

The Ss solved the problem in *ad hoc* groups and as individuals. All of the groups were assigned discussion leaders, only some of whom were trained in techniques of group leadership; the rest were not. The trained leaders, however, knew the "elegant" or experimenter's solution.

Maier's primary conclusion is that a trained leader can improve the quality of the group product. It is limited, of course, to situations where the trained leader knows the experimenter's solution. Under these conditions, Maier also discovered greater individual acceptance of the "elegant" solution. This illustrated the importance of appraising side effects, such as acceptance of a decision, as was done by Lewin (cf. 35).

PRODUCTIVITY

In industrial situations, people work in genuine life settings with real problems. Since the problems are so genuine, the participants are usually motivated, often highly motivated. An example of interaction to influence productivity is Bavelas' study (see 40, pp. 264-266) in which three groups of factory workers met with a psychologist to set a new production

goal for themselves. Their previous production "high" had been 75 units, their previous average was 60 units, and the goal they set was 84 units. This was achieved. Then, they met again and decided to set 95 units as a goal which they failed to realize, but production was stabilized at 87 units. Two other teams of workers serving as controls met with the psychologist but set no goal. They showed no significant variation from their previous average of 60 units.

Coch and French (11) reported a study using as Ss factory workers with an average of eight years of schooling. They were investigating the effects on productivity of three degrees of participation in decision making. All groups had been producing at 60 units before the job change. One group had no participation in making decisions about the job change—management gave them the reasons for the change and answered any questions they raised. Experimental Group I only partially participated in the decision making—they elected a committee which made the decisions. Two other experimental groups participated completely, making all decisions as a group. The decisions pertained to design of the new job, determination of new rates, training methods, etc. The control group dropped, on the average, 10 units per hour, had an exceedingly high rate of turnover, and a slow rate of learning on the new job.

Production in the experimental groups dropped initially but quickly rose to their old levels. In the complete participation group, productivity reached a new "high," 15% higher than previous production rates, the relearning rate was very rapid while turnover was practically nil. Coch and French concluded that

speed of regaining old production levels was directly proportional to the degree of participation in the decision to make the changes. That this was not just a function of the specific personnel involved is emphasized when the control group became the experimental group for a later experiment, showing the same quick recovery and a new production high at the new job.

Marrow (43), president of the company in which Coch and French conducted their studies, reported that union-instituted job changes with money bonuses had failed to change production rates or to get workers to accept job changes. When the program of group decision-making was instituted, a control group that was changed by the customary technique objected bitterly: 17% quit the job, and the rest showed little improvement. The experimental group, on the other hand, as noted earlier reached and then exceeded their old levels with none quitting. Marrow concluded, as did Coch and French, that participation is the key to success in group production.

In 1952 Darley et al. (13) reported a study on group productivity only. The groups were residents of 13 women's cooperative housing units at the University of Minnesota. Each house (or group) contained 7 to 16 students with its own president. The authors stressed that "efforts were made to create an in-group spirit that would characterize residents of the village and give them a feeling of belongingness to the group." The students had been living together in the house for several months, long enough to be considered to have developed a group "tradition." The task was to prepare a "plan for better cooperative living in the village," a first instance of a complex human

relations problem not only in a realistic but "real" situation of actual concern to the Ss. Faculty judges ranked the reports for quality so that cash prizes could be given. In general, studies on productivity have advanced beyond the early use of the group as a social climate or as a means of inducing competition or cooperation. Recent work is moving toward work with people in real situations making a decision basic to their everyday work. This, however, had involved a subtle change from participating in selecting alternatives for a decision, as Lewin's Ss did, to participating in working out the action plan which will give effect to a decision already made, e.g., as it is in the Coch and French (11) study.

Furthermore, productivity experiments have tended not so much to contrast the participation in a group setting with participation in the individual setting, but, rather, to contrast different amounts and kinds of participation where the fullest participation usually was in a group setting. More valid comparisons might have been made if each individual in the control group had been consulted, had had an opportunity to discuss the whole problem, and had made his wishes known. Then the comparison would have been between the effect on productivity of participation as individuals and as a group.

MATHEMATICAL MODELS

As important as the generalizations from experimental data—two changes in methodology over the years and recent attempts to provide structure through mathematical models—are, it is a long step from Gordon's (22) statisticized group, or Allport's (1) noninteracting face-to-face groups to Bavelas' (see 40, pp.

264-266) *ad hoc* groups and Darley's (13) "traditioned" group. Many of the early experiments did not produce results relevant for the problem under investigation. The "statisticized" group gives little evidence about real groups in real situations. As has been indicated, if all the judges are from the same population, the *group products* are sheer reiterations of the value of the Spearman-Brown prophecy formula.

Similarly, noninteracting groups shed more light on the significance of cooperation and competition, with and without an audience, than on the dynamics of grouping for the quality of the final product. Recent work with interacting *ad hoc* groups, and approximations toward the traditioned group, points the direction for estimating the consequences of groups in solving problems in real situations. There is some hope that group dynamics will be understood more fully, a hope more evanescent when grouping was by computation, i.e., without interaction.

A second advance is in the nature of the problem. The older less real tasks of ranking weights or estimating intelligence from photographs is not too crucial a basis for estimating group superiority. Fortunately, the trend is toward human relations problems as exemplified by Maier's parasol assembly line.

Yet, despite the accumulation of evidence suggesting group superiority, the question of the efficiency of the group has not been considered too often. Thorndike has hinted that for some problems, the group may not be as efficient as the individual. Husband (28) gives evidence that in terms of man-hours per correct solution, the group is not as efficient as the individual. The inefficiency in time cost is implicit in

Watson's first study which showed the summated individual superior to the group. There is a practical need to specify the conditions under which the group may be used with the greatest efficiency so as to channel group process to circumstances for which the consequences are commensurate with the time and manpower used. It is on the question of efficiency of the group over the individual that mathematical models can at present play a very instructive role.

In this report we are trying to focus on group productivity in contrast to individual productivity. Another active area of group research for the last 25 years has focussed on the study of group dynamics and group processes without regard to group productivity. Kelly and Thibaut (30) take a strong position on this latter research activity and state that continuing research along these lines "is indicative of major inadequacies in the research field," and that the reason for continuing research of this type "may be suspected to lie in the absence of any good theory about either individual or group processes . . ." (p. 780). While the senior author of this report recognizes the merit of this position, he has the feeling that perhaps a more tenable position is a realization that both approaches have their legitimate place in this broad area of study, and are supplementary instead of contradictory. The one real, though unfortunate, restriction that now exists is that simultaneous experimentation with social dynamics and processes and end products is highly difficult and usually unrewarding. This is probably due to the necessity for concentration on the very basic concepts and minutiae of both areas. The eventual goal is to develop a sizeable body of knowledge in both areas and

perhaps hope for the convergence of this knowledge into one unified and consistent theory. An inherent recognition of this is also contained in Kelly and Thibaut's article in their discussion of the requirements for an understanding of the social processes involved in problem solving. Here again the exploitation of mathematical models might prove useful.

The recent interest in mathematical models is exemplified in papers by Lorge and Solomon (38) and by Hays and Bush (26). Lorge and Solomon provided two mathematical models to reproduce the Shaw data. In their mathematizations they essentially say that the observed better group performance can be explained simply on the basis of individual ability or interaction of individual ability, a conclusion at variance with Shaw's generalization that positive personal interaction is yielding the observed better group performance. Hays and Bush use the Humphreys type learning experiment to mathematically assess performance in groups of three and in individuals. They consider two conceptualizations: one where the group acts as an individual; the second where majority vote in each instance determines the group performance. These two conceptualizations were thought of as boundaries for group performance. An experiment was then run and the results obtained did lie between the two expected group performances. Obviously one importance of mathematization is that it provides the design for a next step in experimentation which might never be revealed solely by examination of data. In the Lorge-Solomon situation some new experimentation was indicated by the results. In the Hays-Bush study new experimentation where *none* had been performed before was indicated by

their two models. An interesting account of these two papers and several others not directly related to problem solving or learning is given by Coleman (12) in a survey of mathematical models in small group behavior.

In general, in the evaluation of the relative quality of the products produced by groups in contrast to those produced by individuals, the group is superior. The superiority of the group, however, all too frequently, is not as great as would be expected from an interactional theory. In many studies, the product of the "best" individual is superior to that of the "best" group. It is quite probable that group solution may have its advantages in stimulating one another for, and in inducing cooperation for, a common solution. Yet, it must be recognized that group procedures may have disadvantages, too. A single member, or a coalition of members, may retard the group by holding out for its kind of solution—a consequent that may reduce the quality of the group product if the solutions so proposed are inadequate or unrealistic.

Obviously, it would be valuable to study group processes to ascertain how members facilitate or inhibit the development of a group product. But, it is just as important to ascertain the quality of product developed by groups in contrast to those of individuals. The researches reviewed in this survey are limited to the experimental assessment of the quality of group and of individual products. Currently, the research literature is in a terminological confusion, generally, in that the group involved in the experimental studies is an *ad hoc* group created by the experimenter for the experimenter's purposes, but that the experimenter generalizes to a traditioned group which has organ-

ized for a common purpose and has been interacting over a considerable period of time. A similar conflict of terms is involved in the definition of the individual. Experimentally, the individual is a person selected at random from some population, but the interpretation suggests that even the best individuals cannot appraise or solve a problem in the round. Since some individuals can, the essential question is to determine how the efficient and able individual compares with the traditioned and able group.

In the researches reviewed, moreover, the range of tasks varies from estimating (e.g., the number of beans in a bottle) to solving genuine prob-

lems (e.g., the productivity of an assembly line). The significance of these tasks in involving individuals and groups does differ and so do the problem-solving processes. Nevertheless, the experimenters tend to interpret the results of all tasks in terms of the relative ability of groups and of individuals to solve novel problems of policy or procedure.

The researches, to date, tend to treat the *ad hoc* group solving a trivial task as a prototype of the fully traditioned group solving a very important problem. The microscopic approach is a prerequisite to understanding the genuine group but is not itself the understanding.

REFERENCES

1. ALLPORT, F. The influence of the group upon association and thought. *J. exp. Psychol.*, 1920, **3**, 152-182.
2. ASCH, M. J. Nondirective teaching in psychology: An experimental study. *Psychol. Monogr.*, 1951, **65**, No. 4 (Whole no. 321).
3. ASCH, S. E., BLOCK, HELEN, & HERTZMAN, M. Studies in the principles of judgments and attitudes: I. Two basic principles of judgment. *J. Psychol.*, 1938, **5**, 219-251.
4. BALES, R. F., MILLS, T. M., ROSEBOROUGH, MARY E., & STRODTBECK, F. L. Channels of communication in small groups. *Amer. sociol. Rev.*, 1951, **16**, 461-468.
5. BANE, C. L. The lecture vs. the class discussion method of college teaching. *Sch. & Soc.*, 1925, **21**, 300-302.
6. BARTON, W. A., JR. The effect of group activity and individual effort in developing ability to solve problems in first year algebra. *Educ. admin. Supervis.*, 1926, **12**, 512-518.
7. BECHTEREV, W., & LANGE, A. Die Ergebnisse des Experiments auf dem Gebiete der kollektiven Reflexologie. *Zeit. angew. Psychol.*, 1924, **24**, 305-344.
8. BRUCE, R. S. Group judgments in the fields of lifted weights and visual discrimination. *J. Psychol.*, 1935-36, **1**, 117-121.
9. BURTT, H. E. Sex differences in the effect of discrimination. *J. exp. Psychol.*, 1920, **3**, 390-395.
10. CARTER, L., HAYTHORN, W., LANZETTA, J., & MAIROWITZ, BEATRICE. The relation of categorizations and ratings in the observation of group behavior. *Hum. Relat.*, 1951, **4**, 239-254.
11. COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512-532.
12. COLEMAN, J. S. A survey of mathematical models of small group behaviour. New York: Bureau of Applied Social Research, 1956. (Behavioral Models Project Tech Rep. No. 10.)
13. DARLEY, J., GROSS, N., & MARTIN, W. Studies of group behavior: Factors associated with the productivity of groups. *J. appl. Psychol.*, 1952, **36**, 396-403.
14. DASHIELL, J. F. An experimental analysis of some group effects. *J. abnorm. soc. Psychol.*, 1930, **25**, 190-199.
15. DASHIELL, J. F. Experimental studies of the influence of social situations on the behavior of individual human adults. In C. Murchison (Ed.), *Handbook of social psychology*. Worcester: Clark Univer. Press, 1935, Pp. 1097-1158.
16. EYSENCK, H. J. The validity of judgments as a function of number of judges. *J. exp. Psychol.*, 1939, **25**, 650-654.
17. EYSENCK, H. J. The validity and reliability of group judgments. *J. exp. Psychol.*, 1941, **29**, 427-434.
18. FARNSWORTH, P., & WILLIAMS, W. F.

- The accuracy of the median and the mean of a group of judges. *J. soc. Psychol.*, 1936, **7**, 237-239.
19. GIBB, J. R. The effects of group size and of threat reduction upon creativity in a problem solving situation. *Amer. Psychologist*, 1951, **6**, 324. (Abstract)
 20. GORDON, KATE. A study of aesthetic judgments. *J. exp. Psychol.*, 1923, **6**, 36-42.
 21. GORDON, KATE. Group judgments in the field of lifted weights. *J. exp. Psychol.*, 1924, **7**, 389-400.
 22. GORDON, KATE. Further observations on group judgments of lifted weights. *J. Psychol.*, 1935-36, **1**, 105-115.
 23. GURNEE, H. A comparison of collective and individual judgments of fact. *J. exp. Psychol.*, 1937, **21**, 106-112.
 24. GURNEE, H. Maze learning in the collective situation. *J. Psychol.*, 1937, **3**, 437-444.
 25. GURNEE, H. The effect of collective learning upon the individual participants. *J. abnorm. soc. Psychol.*, 1939, **34**, 529-532.
 26. HAYS, D. G., & BUSH, R. R. A study of group acting. *Amer. sociol. Rev.*, 1954, **19**, 693-701.
 27. HILGARD, E. R., MAGARET, G. A., & SAIT, E. M. Level of aspiration as affected by relative standing in an experimental social group. *J. exp. Psychol.*, 1940, **27**, 411-421.
 28. HUSBAND, R. W. Cooperative versus solitary problem solution. *J. soc. Psychol.*, 1940, **11**, 405-409.
 29. JENNESS, A. The role of discussion in changing opinion regarding a matter of fact. *J. abnorm. soc. Psychol.*, 1932, **27**, 279-296.
 30. KELLY, H. H., & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 735-785.
 31. KLUGMAN, S. F. Group judgment for familiar and unfamiliar materials. *J. genet. Psychol.*, 1945, **32**, 103-110.
 32. KLUGMAN, S. F. Group and individual judgments for anticipated events. *J. soc. Psychol.*, 1947, **26**, 21-33.
 33. KNIGHT, H. C. A comparison of the reliability of group and individual judgments. Unpublished master's thesis, Columbia Univ., 1921.
 34. LEVINE, J., & BUTLER, J. Lecture versus group decision in changing behavior. *J. appl. Psychol.*, 1952, **36**, 29-33.
 35. LEWIN, K. Group decisions and social change. In T. Newcomb & E. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947.
 36. LORGE, I., DAVITZ, J., FOX, D., & HERROLD, K. Evaluation of instruction in staff action and decision-making. *USAF Hum. Resource Res. Inst. Tech. Rep.*, 1953, No. 16.
 37. LORGE, I., DAVITZ, J., FOX, D., HERROLD, K., & WELTZ, PAULA. Studies contrasting the quality of products by individuals and by groups. Unpublished manuscript, Institute of Psychological Research, Teachers College, Columbia Univ.
 38. LORGE, I., & SOLOMON, H. Two models of group behavior in the solution of Eureka-type problems. *Psychometrika*, 1955, **20**, 139-148.
 39. LORGE, I., AIKMAN, L., MOSS, GILDA, SPIEGEL, J., & TUCKMAN, J. Solutions by teams and by individuals to a field problem at different levels of reality. *J. educ. Psychol.*, 1955, **46**, 17-24.
 40. MAIER, N. R. F. *Psychology in Industry*, Boston: Houghton Mifflin, 1946.
 41. MAIER, N. R. F. The quality of group decisions as influenced by the discussion leader. *Hum. Relat.*, 1950, **3**, 155-174.
 42. MARQUART, DOROTHY I. Group problem solving. *J. soc. Psychol.*, 1955, **41**, 103-113.
 43. MARROW, A. Group dynamics in industry: Implications for guidance and personnel workers. *Occupations*, 1948, **26**, 472-476.
 44. MARSTON, W. M. Studies in testimony. *J. crim. Law. Criminol.*, 1924, **15**, 5-31.
 45. MOORE, O. K., & ANDERSON, SCARVIA. Search behavior in individual and group problem solving. *Amer. sociol. rev.*, 1954, **19**, 702-714.
 46. MUKERJI, N. F. Investigation of ability to work in groups and in isolation. *Brit. J. Psychol.*, 1940, **30**, 352-356.
 47. PERLMUTTER, H. V. Group memory of meaningful material. *J. Psychol.*, 1953, **35**, 361-370.
 48. PERLMUTTER, H. V., & DE MONTMOLLIN, GERMAINE. Group learning of nonsense syllables. *J. abnorm. soc. Psychol.*, 1952, **47**, 762-769.
 49. PRESTON, M. Note on the reliability and validity of the group judgment. *J. exp. Psychol.*, 1938, **22**, 462-471.
 50. RADKE, MARIAN, & KLISURICH, D. Experiments in changing food habits. *J. Amer. diet. Ass.*, 1947, **23**, 403-409.
 1. ROBINSON, K. F. An experimental study

- of the effects of group discussion upon the social attitudes of college students. *Speech Monogr.*, 1941, **8**, 34-57.
52. RYAN, G. An experiment in class instruction versus individual study at college level. Unpublished doctoral dissertation, Johns Hopkins Univer., 1932.
 53. SCHONBAR, ROSALEA. The interaction of observer pairs in judging. *Arch. Psychol.*, N. Y., 1945, 299.
 54. SENGUPTA, N. N. & SINKA, C. P. N. Mental work in isolation and in group. *Indian J. Psychol.*, 1926, **1**, 106-109.
 55. SHAW, M. E. Comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.*, 1932, **44**, 491-504.
 56. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, N. Y. 1935, 187.
 57. SIMS, V. M. The relative influence of two types of motivation on improvement. *J. educ. Psychol.*, 1928, **19**, 480-484.
 58. SMITH, B. The validity and reliability of group judgments. *J. exp. Psychol.*, 1941, **29**, 420-426.
 59. SMITH, M. Group judgments in the field of personality traits. *J. exp. Psychol.* 1931, **14**, 562-565.
 60. SOUTH, E. B. Some psychological aspects of committee work. *J. appl. Psychol.*, 1927, **11**, 348-368.
 61. SPENCE, R. B. Lecture and class discussion in teaching educational psychology. *J. educ. Psychol.*, 1928, **19**, 454-462.
 62. STROOP, J. B. Is the judgment of the group better than that of the average member of the group. *J. exp. Psychol.*, 1932, **15**, 550-560.
 63. TAYLOR, D. W., & FAUST, W. L. Twenty questions: Efficiency in problem solving as a function of size of group. *J. exp. Psychol.*, 1952, **44**, 360-368.
 64. THIE, T. W. The efficiency of the group method. *English J.*, 1925, **14**, 134-137.
 65. THORNDIKE, R. L. The effects of discussion upon the correctness of group decision when the factor of majority influence is allowed for. *J. soc. Psychol.*, 1938, **9**, 343-362.
 66. THORNDIKE, R. L. On what type of task will a group do well? *J. abnorm. spec. Psychol.*, 1938, **33**, 408-412.
 67. TIMMONS, W. M. Decisions and attitudes as outcomes of the discussion of a social problem. Teachers Coll., Contrib. Educ., No. 777, Columbia Univer., Bureau of Publications, 1939.
 68. TIMMONS, W. M. Can the product superiority of discussors be attributed to averaging and majority influences? *J. soc. Psychol.*, 1942, **15**, 23-32.
 69. WAPNER, S., & ALPER, THELMA G. The effect of an audience on behavior in a choice situation. *J. abnorm. soc. Psychol.*, 1952, **47**, 222-229.
 70. WATSON, G. B. Do groups think more efficiently than individuals? *J. abnorm. soc. Psychol.*, 1928, **23**, 328-336.
 71. WATSON, G. B. A comparison of group and individual performances at certain intellectual tasks. *Proc., Ninth intern. Congr. Psychol.*, 1929, 743.
 72. WYATT, S., FROST, L., & STOCK, F. G. L. Incentives in repetitive work. *Med. Res. Coun. Rep.*, 69. London: H. M. Stationery Office, 1934.
 73. ZELENY, L. D. Teaching sociology by a discussion group method. *Sociol. Soc. Res.*, 1927, 162-172.
 74. ZELENY, L. D. Experimental appraisal of a group learning plan. *J. educ. Res.*, 1940, **34**, 37-42.

Received January 15, 1958.

INDICATORS OF PERCEPTION: I. SUBLIMINAL PERCEPTION,
SUBCEPTION, UNCONSCIOUS PERCEPTION: AN ANALYSIS
IN TERMS OF PSYCHOPHYSICAL
INDICATOR METHODOLOGY¹

ISRAEL GOLDIAMOND

Southern Illinois University

Public concern is currently being expressed over commercial exploitation of subliminal perception. This use of subliminal perception is probably related to the impetus given studies in unconscious perception by work in perceptual defense and subception, a neologism from *subliminal perception*. Two reviews of the literature have recently appeared. One includes experiments in subliminal perception (1), and the other in subception and perceptual defense (92). Both reviews systematize the literature, but neither pays sufficient attention to the relationship between the psychophysical methods employed and the data collected. No recent review or discussion (cf. 193) refers to current developments in psychophysics which are a major breakthrough in this area, and which necessitate reformulation of many concepts in perception. In contrast to preceding reviews, this review will concentrate systematically on psychophysical indicator methodology. The major point of this paper will be that much of the controversy in the area of unconscious perception is peripheral to the central perceptual problem of psychophysical procedures utilized.

¹ The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, and was made possible by grants provided by the National Science Foundation, to whom the author is greatly indebted for generous support.

The focus of this review will be experiments in which *E* seeks to obtain thresholds or related perceptual measures. Since the discussion will be concerned with methodology, concentration will be on experimental investigations rather than on discussions or historical accounts; such antecedents may be found in other reviews (e.g., 32, 34). A further limitation is given in the title. This discussion will consider those experiments where the perceptual measures obtained are related to unconscious processes by *E* or others in the field, either in affirmation or negation. Additional material will be presented from research in which this is not an issue, but which is relevant to the discussion. A later paper will extend the methodological analysis to pertinent areas of perception not covered.

No position will be taken on the issue of defining perception as: (a) a concept which relates certain sets of operations and logical relations, (b) a sensation related to physiological variables and often implying a subjective experience, (c) a subjective experience. This review, however, will be concerned with all three definitions since it will consider systematically what is procedurally common to all definitions, namely, a response indicator.

By *indicator* (14) is meant that class of responses which contributes the response ingredient (often called the dependent variable) to the definition of perception which *E* accepts.

In addition to specified sets of responses or operations (cf. 67), perception experiments are also *defined* by certain experimental procedures, logical relations, historical antecedents. All of these will be subsumed under *methodology*. Perception can be considered simply as a concept (or schema)² methodologically defined in reference to *O*, and measured by the indicator. Use of the indicator may be extended to measure some sensation or process which may imply a subjective experience, or to measure some state of mind, such as awareness. Although verbally the most cumbersome, the schematic usage of the indicator is scientifically the simplest, since, no matter what extensions are made, the adequacy of the response to indicate perception will depend upon the methodological adequacy of the experiments involved. If, for example, *E* requires *S* to press a button every time a light goes on, and the button requires 25 pounds of pressure, the response will not continue for long. The *E* should not then be expected to report either that perception (concept) had been attenuated, or that visual sensation had diminished, or that awareness of the stimulus had fallen off. There would be agreement among all describers that the procedures employed had invalidated the indicator, and for precisely the same reason that a test is invalidated: the indicator is admitting variance from extraneous sources. In the button case cited, it

is a historically demonstrated fatigue variable which is the extraneous source. Had this variable not been previously isolated, there might be argument concerning perception. In other cases, the indicator might be invalidated through the use of logic or operations not covered by the methodological definition. Hence, the methodology which defines perception minimally is basic to all definitions and inferences having added meanings as well.

I. INDICATORS OF PERCEPTION AND THE SUBLIMINAL PERCEPTION PARADIGM

In contrast to the usual perception experiment where only one indicator is investigated, all experiments in subliminal perception deal with differences between two indicators of perception, at related stimulus measures.

Of the two indicators of perception involved in subliminal perception, one is interpreted as an indicator of absence of awareness, that is, absence of awareness of the perception of a stimulus. The other indicator is interpreted to indicate perception or stimulus discrimination of the same stimulus. Subliminal perception, discrimination without awareness, and the subliminal effect (the term to be used in this discussion), refer to the presence of both indicators at the same stimulus magnitude. The indicators can be separated in time. For example, a stimulus magnitude is found which relates to reports of no awareness; at this magnitude or lower *S* later correctly identifies geometric forms. This difference will be called the *discrepancy* effect. The indicators can be contemporaneous. For example, at a given stimulus magnitude *S* reports no awareness and also makes a correct identification. This

² Actually, the term *construct*, alone or in combination as *construct-perception*, is preferable to *concept* or *schema* on a variety of grounds, among them etymological, and conveys the meaning intended. Unfortunately, its meaning has tended to become restricted, among psychologists, at least, to indicate the usage implied in the classical differentiation between *hypothetical construct* and *intervening variable*.

will be called the *asynchrony* effect.

The foregoing constitute the paradigms for subliminal perception experiments. All involve the first indicator discussed. The second indicator varies, and forms the basis for the classification to be used in this discussion.

The category with the greatest number of studies involves an accuracy indicator. The *S* is scored for correctly identifying a geometric figure, letter, number, and so on. Asynchrony effects were obtained by Baker (5) using judgments of rotation of visual figures, by Collier (32) for visual and tactile forms, by Coover (33) for letters and numbers, by Landis and Vinacke (102) for colors, by Peirce and Jastrow (132) for increase or decrease in tactile pressure, by Vinacke (186) for colors, and by Williams (194) for geometric forms. Discrepancy effects were obtained by Binet (8) using letters, and also by Coover (34), by King, Landis, and Zubin (94) using geometric forms, by Miller (121) using E.S.P. forms, and other visual forms (122), by Pillai (133) for letters voiced and shown, by Sidis (155) for letters and numbers whispered and shown, and similarly by Stroh, Shaw, and Washburn (173).

In contrast to the foregoing studies, which report subliminal effects, conflicting results have characterized use of an illusion indicator. Dunlap (46) interpreted results to faint arrows of the Müller-Lyer illusion as indicating no awareness, and obtained size judgments in accord with the illusion. Titchener and Pyle (182) could not confirm this, and were supported by Manro and Washburn (118), 8 of whose *Ss* were unsuccessful, while 2 did obtain an illusion. Hollingworth (81) goes into very little detail concerning the method

used by a student to get illusions in 17 of 20 *Ss*. Bressler reported the illusion "increases gradually as the strength of the stimulus increases" (20, p. 250).

A tone previously reported unheard is shut off; this change is then reported by *S* in experiments by Dunlap (47), Dunlap and Wells (48), and Jastrow (91). Colored after-images to color reported unseen were obtained by Fernald (59), challenged by Titchener and Pyle (183), and reaffirmed by Ferree and Rand "under the right conditions" (60, p. 196). These results were again obtained by the same authors (61), and by Newhall and Dodge, who found that "thresholds for after-images were lower than for stimulus color" (128, p. 8). De Laski (43) found thresholds for cutaneous form were lower than the two-point threshold, as did Friedline (66).

In "conditioning without awareness," responses to visual or auditory stimuli are interpreted to indicate no awareness, and a conditioned stimulus is presented at this magnitude or lower. The unconditioned stimulus is supraliminal. Baker (6) reported pupillary conditioning to a subliminal tone, the unconditioned stimulus being light. These results could not be replicated by Hilgard, Miller, and Ohlson (77), nor by Wedell, Taylor, and Skolnick (188). Indeed, Steckle (169) and Steckle and Renshaw (170) had indicated difficulties in assaying even the supraliminal conditioning of this reflex previously reported by Cason (25), and Hudgins (88). A recent (1956) failure to obtain supraliminal pupillary conditioning is reported by Crasilneck and McCranie (38). Wilcott (192) has also reported negative results with subliminal stimuli, as have Cason and Katcher (26). Taylor (178) has re-

ported positive results. Subception studies involve different procedures and will be discussed separately.

Adams (1) presents a chart listing the various experiments with a positive sign or a negative sign for conclusions pro and contra the subliminal effect. This recalls an earlier summary which reported a score of "three to one against Titchener and Pyle, with one tie" (121, p. 564), and implies that scientific issues can be settled by nose-counts, with 10 poorly controlled studies reaching one conclusion outvoting a well-designed dissenting vote. The studies discussed vary widely in methods and controls. Binet (8) apparently bases his non-awareness indicator on *one* descending series per *S*. A hysteric walks backward until he reports he can no longer read the largest letter on a chart; he then writes smaller letters. The Titchener and Pyle (182) repetition of Dunlap's study (46) provided controls such as having the variable on both sides, instead of one, light adaptation, practice, better equipment, and better analytic procedures. These two studies, at least, cannot be equated. "Whispers" are used (155, 173), arrows are pencilled-in (118), there are time-lapses between each of the two indicators used (94, 121, 122, 133, 155, 173). A more detailed discussion of some individual experiments is presented by Wilcott (193). Full psychophysical functions are seldom obtained. On the other hand, Davis' study (42) involving muscle action potentials to subliminal auditory stimulation is straightforward, as are many of the others cited.

Semantic and Accuracy Indicators

Use of one of the responses in subliminal perception experiments to indicate awareness has historical precedent in early psychophysical studies

where Yes, No, I see are assumed to have stimulus-related experiential referents. These referents are considered different from those attached to guessing a triangle correctly or where muscle action potential is related to auditory stimulation. Because of the historical application of such semantic referents to Yes-No and similar responses, this class of indicators will be called *semantic indicators*.

Needless to say, the continued use of such indicators has not been predicated upon their correspondence to common sense semantics, but rather upon their methodological fit, that is, upon their long history of continued lawful relations with other variables under specified conditions. It has accordingly been argued that the semantic referents of the symbolic responses which constitute this indicator are irrelevant to it, and that the responses can be considered simply as responses with properties such as frequency, rate, amplitude, latency, duration. Regarding this argument, it should be noted that the term, *semantic indicator*, is used here in a neutral manner. Use of this term should not be construed as implying a semantic referent for this indicator. It is the name of a class of responses which can be so interpreted and which also need not be so interpreted.³

³ A behaviorist position holds that emotions, drives, states, sensations can be defined in terms of differing sets of logical relations and empirical relations between observable changes in environment and organism. The *name* that is given to each of these often turns out to be the name of the experience involved (cf. 58), e.g., love, hunger, anxiety, vision. Using such a term (or the term semantic indicator) does not imply that the subjective experience which is the common sense referent of the term is being discussed. Someone wishing to use the sets discussed to indicate an experience may do so. But when he wishes to

The indicator paired with the semantic indicator varies in the studies mentioned. Some of the miscellaneous studies in cutaneous differences, color phenomena, and auditory on-off phenomena have been cited in discussions of subliminal perception, but can be considered peripheral to the discussion since they can be interpreted to indicate that within the same modality, there is differential sensitivity to classes of stimuli. There are, for example, threshold differences for blue and yellow, motionless stimuli and those in motion, high and low pitch, and so on. The Müller-Lyer illusion issue is by no means settled, and the conditioning studies would seem negative. The main body of data supporting the subliminal effect comes from studies in "discrimination without awareness" (121), where the semantic indicator is coupled with an accuracy indicator. Since two different indicators are involved, it is conceivable that differences in results may be related to systematic differences between the indicators themselves, such as types of scores taken, response control, correction procedures, and other such differences which are unrelated to the substantive issues raised. Accordingly, a systematic comparison follows.

Response and semantic referent of response. Accuracy responses usually designate the stimulus in some way. They identify it (e.g., stimulus is Triangle, Circle, 5, E), locate it

spatially (e.g., stimulus Up or Down, break in circle faces East) or temporally (e.g., stimulus was presented in first interval), count it (e.g., 4 dots). Semantic responses usually do not have this stimulus-designation feature. Rather, they seem to have subjective referents such as Yes, Confident, Seen, or terms which stand for them (e.g., use 2 for Yes and 1 for No; underline once if Certain and twice if Doubtful). This difference per se is not too useful for methodological classification. For example, Yes is an accuracy indicator when *E* presents a signal during one of *n* intervals, the other intervals containing only noise, and *S* must state Yes during one of the intervals, and is scored for congruence between Yes and signal. On the other hand, Down can be a semantic indicator. *E* can present a column of alternating squares and circles, with colors also alternating. Using phi-phenomenon procedures, the colors move up, but the forms move down. Although *S* is forced to choose between movement Up or Down, if each of these is an equally valid mode of response (for a color-form categorization), this is not an accuracy indicator, despite the fact that discrimination is forced in terms of specified stimulus spatial relationships. The features which distinguish semantic and accuracy indicators and which are methodologically meaningful follow.

1. Indicator score. The accuracy score is not *S*'s responses, but the accuracy of these. Responses are converted to congruence between responses and *E*'s answer-grid sheet. The semantic score utilizes a measure of the response itself.

2. Control over score. Since congruence defines accuracy, *E* exerts considerable control over the score by his control of the answer sheet. This

conduct a test, it is the set he must test, since this can be observed and agreed upon by others, but the experiences cannot. For the present discussion, if we can agree on proper tests of variables affecting the sets under discussion, we can be tolerant to differences involving subjective or operational meanings of terms.

is not the case in the semantic indicator.

3. Correction of indicator score.

(a) Chance congruence. The accuracy congruence discussed can occur by chance alone. Correction involves consideration of the number of alternatives on *E*'s sheet. The term, *accurate*, as used here, implies such a correction. There is no comparable a priori method of correction for semantic indicators. The probability of Yes in Yes-No is not 1/2.

(b) Response bias: general and accuracy. *Bias* is used here in the sampling sense, namely, that not all responses will have an equal probability of appearing in the response sample. It might be said that a psychophysical experiment is concerned with response biases of an observer which are related to the stimulus. Irrelevant bias must be controlled. In accuracy indicators, two procedures are available. One involves presetting sequences on the answer sheet so that congruences obtained through response bias are balanced by incongruences so obtained. An *S* who tends to respond left will be spuriously accurate when the answer is left, but his score will be attenuated by answer sheet rights put in for this purpose by *E*. A second procedure involves analysis of the response record for this same purpose.

(c) Positive semantic bias. Similar correction here is complicated by *E*'s comparative lack of control over the score. To control biases to respond Yes, catch stimuli such as blanks are often thrown in (the reference here is to the use of blanks for *score* correction, rather than for disciplinary correction of *S* when he says Yes to a blank). From false positives here (coupling of Yes and blank), a correction factor is drawn

for application to Yes responses whose "falsity" can not be readily determined, that is stating Yes when a stimulus is presented. If both accuracy and semantic responses are made to the same stimulus presentation, false positives can be obtained without using catch stimuli, that is, when *S* trips himself by coupling Yes with an inaccurate response. The assumptions upon which these correction procedures rest have recently been questioned. This issue will be discussed in the section on decision processes.

(d) Negative response bias. If response bias can lead to false positives, it must logically be capable of leading to false negatives. If correction procedures are used for one, they should logically be used for the other. As a rule, no correction procedures are used for false negatives, and the *existence of false negatives defines the subliminal perception effect*: *S* says No and is accurate. The subliminal effect is accordingly contaminated by a procedural and logical inconsistency. Further, this one-sided correction decreases the number of Yesses and can therefore in and of itself produce a discrepancy or asynchrony effect, unless there is corresponding correction of false negatives.⁴

⁴ For example, a task will be assumed in which accuracy by chance is highly improbable (signal in 1 of 1,000 positions). The *S* says Yes 10 times and is accurate 7 times, and says No 10 times and locates signal accurately 3 times. This gives the following small table:

		S's location was:	
		Inaccurate	Accurate
S reported:	Yes	3	7
	No	7	3

One way of interpreting this table is to state the *S*'s responses are quite valid, but he is just as likely to "misuse" Yes as he is No. Another interpretation involves the establishment of a criterion for reporting Yes such that for inputs

4. Decision avoidance. Although in both indicators a response may be required, that is, forced, experimental conditions are often such that the negative semantic response can be used simply to avoid making a decision. Culturally, the negative response is preferred to the positive response for equivocal situations. This amounts to a cultural bias toward a response for which no correction is made.

5. Cross-experimental validation. The logic of the accuracy indicator can be independent of the responses used. For example, *S* is required to respond Triangle, Circle, or Square when one of these is presented. The better he can differentiate signal from background, and the better he can relate signal configurations to the response, then the greater the prob-

ability of congruence. This holds whether the responses are geometric forms, or letters, figures, and so on, although there are some major differences between certain procedures. Nuances in instructions will not usually affect this logic (although complexity of the task may make a difference). The semantic indicator, on the other hand, is considerably more dependent on the response and the instructions which often give it a referent. Some semantic responses used have been: great, slight, no confidence (5); stimulus, something, nothing (34); something, nothing (186); clearly, doubtful but something, guess (194). Instructions vary from "Report when you can no longer see the figure" (94, p. 61), to "be very careful in the matter of giving immediate expression to even the slightest confidence" (5, p. 89). Systematic investigation of the effects of such differences is usually not involved in the studies, and it becomes difficult to compare data and to discuss semantic results obtained independently of specific responses and instructions utilized.

The following sections concern systematic investigations of these two indicators.

Semantic and Accuracy Indicators in Recent Psychophysical Research; Validity Differences

Recent work in psychophysics bearing directly on the two indicators discussed may affect conceptualization and procedures in perception and mental test theory.

The history of psychophysics can be interpreted as a search for methods and indicators which control invalidating extraneous sources of variance. The term, *methods*, usually brings to mind procedures such as the Constant Methods, Method of

difficult to discriminate, exclusion of false positives and exclusion of true positives covary positively, with a high criterion leading to high exclusion rates (less Yes responses of each kind), and low criterion to low exclusion rates. Decision theory, to be discussed shortly is concerned with such decisions, and might interpret this data in terms of *S* maximizing response value by using a criterion producing 3-7 ratios.

A third way is to look at the accurate column only and note that in regard to *S*'s accurate locations of the signal, 30% of these occurred without awareness! This is then "explained" by reference to unconscious processes in subliminal perception.

Any correction for positive or chance bias of the type discussed would only increase the subliminal effect. In this example, placing the signal in 1 of 1,000 possible positions eliminated, for practical purposes, the necessity for a chance correction.

An ingenious dissertation by Zeitlin (197) uses as its central argument a similar treatment of perceptual defense, sensitization, and autism. Zeitlin argues that these effects are derivable from interpretations of the data which concentrate only on specified entries in a 2x2 table, and ignoring the total table. Discussion of this paper will be deferred until the section on these effects.

Limits, Average Error, and so on. There has been considerable research in the procedural issues raised by these, but comparatively little systematic research in the procedural issues raised by indicator differences.⁵

An outstanding exception is a monograph (14) and series of studies (11, 12, 13) by Blackwell. These report experimental comparison of semantic and accuracy indicators, called *phenomenal report* and *forced-choice*. The phenomenal report indicator is a semantic indicator standardized to Yes-No. The forced-choice indicator is an accuracy indicator standardized so that the task involves differentiating a signal-plus-noise from noise alone. A light increment is added to a uniformly illuminated screen, and appears in one of several positions (spatial forced-choice), or the increment is presented during one of several intervals (temporal forced-choice). The better *S* can differentiate signal-noise from noise, the greater the probability of the congruence called accuracy. This procedure is simpler than those where differentiating one geometric form from others is the task, since the latter may involve not only signal-noise differentiation, but at least patterning of signals, differentiating one pattern from another, and a wider array of responses. Blackwell's monograph reports systematic experimental variation of the following variables related to indicators, methods, and extraneous variances (groups within each variable are given in parentheses): indicator (Yes-No, forced-choice spatial, forced-choice tem-

poral), number of forced-choice alternatives [2, 4], stimulus magnitude orders (random, ascending, descending, blocked), block size [1, 20, 40], number of magnitudes, threshold relations (below, above and below), motivational and set variables (e.g., variations in instructions and questions; group which are "relaxed," "motivated"), knowledge of correctness. Catch stimuli were employed and varied systematically. The *Ss* were run for several sessions, with only one procedure used at a session.

Comparing semantic and accuracy indicators, Blackwell found that the Yes-No indicator was greatly amenable to influence from variables extraneous to sensory discrimination. If these constitute perceptual error variance, this indicator should, according to test theory, exhibit less session-to-session reliability, and less intrasession reliability than the forced-choice indicator. This is precisely what was found; the "data suggest that forced choice rather than phenomenal report should be used in routine psychological measurement" (14, p. 118). If this indicator is including extraneous variance, it should also yield higher thresholds than the forced-choice indicator, for the same reason that a voltmeter with steel filings in its bearings will be less sensitive than a clean voltmeter. It was reported by Blackwell "that forced-choice thresholds were significantly smaller than corresponding thresholds obtained with phenomenal report" (14, p. 199). The Yes-No indicator had considerably less "apparent validity" than forced-choice.

The discrepancies obtained are entirely in keeping with the subliminal effect. It can be asserted that "a subject can discriminate intensities too low for him to be aware of them. This fact is probably best explained

⁵ Notable among systematic studies on indicator methodology have been those concerned with the response in the P.S.E., and those studies which led to the elimination of middle categories such as the Doubtful in Yes, Doubtful, No.

in terms of levels of operation of the nervous system, some involving consciousness and others not—these levels having different thresholds" (123, p. 265). An alternative statement involves two indicators of perception, one freely admitting extraneous variance, and the other not.

An experiment by Goldiamond (70) compares semantic and accuracy indicators for serial effect, an issue which has recently received considerable attention. Serial effect refers to the covariation of responses in a series with preceding responses. Such psychophysical responses are accordingly admitting variance extraneous to stimulus measurement. Goldiamond found that the probability of saying Yes was not only a positive function of judgment intensity, but also of the stimulus intensity *preceding* the one at which judgment was made, that is, *S* was more likely to say Yes if the preceding stimulus had been high than if it had been low. Since high stimulus magnitudes are associated with high frequencies of responses made in their presence, the increase in Yes which followed might relate either to response effects or stimulus effects. A paired accuracy indicator did not covary with the same preceding stimulus magnitude, eliminating stimulus involvement, and assigning the effect to the response. Congruence with *E*'s score sheet defines this indicator. The *E*'s score sheet not having a serial effect, any response bias of this kind would not enter into the indicator as a serial effect.

These two indicators were further compared by the same author (69) whose *S*s gave both phenomenal report and spatial forced-choice responses to a triangle varied in position and intensity. The *S*s were divided into 6 groups, with visibility

reported as (a) Yes-No, (b) 2-1-0, (c) 7-6- . . . 0, and knowledge of accuracy (a) given immediately after each set of responses, and (b) not given. Only one session was run per *S*. Accuracy curves were similar for all groups; the corrected phenomenal report curves for each group in order of decreasing response frequency (and increasing subliminal effect) were: 7-0 without knowledge (minimal asynchrony), 7-0 with, 210 without, 210 with, Y-N without, Y-N with (35% asynchrony at threshold). This suggests that subliminal effects ranging from zero on up can be wilfully obtained by *E* through proper experimental manipulation.

These data were interpreted in terms of stimulus generalization: an *S* trained to respond No to the absence of a stimulus, S^A , and Yes to its presence, S^D , may upon the presentation of a faint but *discriminable* stimulus be more likely to respond No than Yes, since the faint stimulus is closer to S^A than it is to S^D . Permitting intermediate response would lead to more Yes responses. The depressant effects of knowledge of accuracy were related to elaboration of a 2×2 Utility Table (semantic indicator No-Yes by forced-choice location Inaccurate-Accurate). The combination Yes-Inaccurate (stimulus reported seen but in a location where it was not presented) attached aversive consequences (hallucination, poor judgment, lie) to responding Yes when S^D and S^A were close (72).⁶

⁶ Another interpretation that deserves mention is the linguistic-experiential one. This states that language classifies our experiences, a discrete language such as ours doing so discretely. The *S* is told to label a bright stimulus experience with Yes, and the absence of such experience with No. He then gets a faint but discernable stimulus experience. If he must classify this experience by some response, since it is closer to the experi-

A monograph by Smith and Wilson (159) reports variation of the discreteness of No (reported as (a) Didn't hear but guess, and (b) No) as well as of Yes (reported as (a) Certain heard, and (b) Think). Such "liberal" Ss were contrasted to "conservatives" (Yes-No). Liberals gave more positive responses than conservatives. Had their higher curves merely been an artifact of increased positive response bias, as their high false positive rate suggested, then correction for false positives should have eliminated the superiority of this group, but "the data consistently failed, by a wide margin, to meet this assumption" (159, p. 34). Superiority in signal detection was related to increased false positives. To the usual model of the threshold, which depicts the probability that *S* will receive information at a given pres-

ence labeled No than it is to the Yes experience, he may say No. But he will be able to locate the stimulus accurately since he is experiencing it. If intermediate categories of response are introduced, he may classify the faint experience with a 2 or 1, and the brighter ones with higher numbers. Eriksen (54) relates discreteness in language to the subception effect; the study of language as an instrument classifying experience has recently received considerable attention from anthropologists such as Whorf (190), who would hold that it alters experience, and Hoijer (80), who presents differing views.

The three interpretations given here, decision theory, operant conditioning, and experiential-linguistic, are in agreement in that they differentiate perception from the response (cf. 67). Naturally, they differ on other issues, but on this issue they support the statement made earlier that if there is agreement on proper tests of variables, there can be tolerance to differences of added meanings. The statement by Solomon and Howes (164) that "any variable that is a general property of linguistic responses must also be a property of any perceptual concept that is based upon those responses" (p. 257) would be in similar agreement if altered to refer to perceptual response, rather than perceptual concept, which differs from it.

entation as related to distributions of signal-noise ratios, a third dimension was added. This involved the superimposition of a distribution of cut-off points for *reporting* having received information. "Variations of the observer's attitude toward reporting moves a cut-off line up or down the dimension of subjective intensity. Any signal (or blank) exceeding this cut-off point is reported" (159, p. 34), the authors conclude.

Decision Theory and Signal Detection

Work by Blackwell which systematically investigates differences between indicators, and related research in decision theory and signal detection constitute a major breakthrough in psychophysics and perception. This breakthrough is both methodological and theoretical, and bids to supply new applications of psychophysical techniques as well as insights into new areas. This discussion will be concerned with extensions to unconscious perception.

Decision theory is concerned, in part, with the cut-off issue raised by Smith & Wilson, and with differences between phenomenal report and forced-choice indicators. It questions certain assumptions held in most psychophysical work and pertinent to subliminal perception. Among these are (a) the existence of "such a thing as a sensory threshold . . . (rather than) continuous reception of information" (175, p. 403), and, in this context, "that if some threshold of neural activity is exceeded, phenomenal seeing results" (177, p. 402), and (b) that false positives represent guesses or are not related to signal detection.

Although a distinguishing characteristic of the theory is its mathematical nature, a verbal presentation, minimally necessary for understanding its relation to subliminal percep-

tion, will be made; Swets, Tanner, and Birdsall (174, p. 56) explicitly indicate a relationship. Although the presentation in this discussion is based on work with simple stimulus dimensions, Egan (176) has reported extension of the theory to experiments involving two-way communication of speech. Substantive findings (that is, results of experimentation in vision and hearing) using these formulations will not be discussed.

Most simply, the theory starts with the notion of a signal presented against a background of noise, for example, a tone, or a light increment against a uniformly illuminated screen. Output of (or sensitivity to) signal-plus-noise and noise alone will vary, each around a mean intended as their values by E . A distribution can be assigned to signal-noise, and to noise, with amount of overlap dependent on the variance and closeness of the means. In this area of obvious overlap (and to lesser extent elsewhere), each of these produce similar inputs, and S can call a signal presentation noise, and a noise presentation signal. S 's task is to decide if a given "observation is more representative of N (noise) or $S+N$ (signal superimposed on background noise). His task is, in fact the testing of a statistical hypothesis" (177, p. 403). For Yes-No curves, he can "establish a cut-off point such that any measure which exceeds that cut-off is in the criterion" (177, p. 403) and in so doing, engages in the risks attached to Type I and Type II statistical decisions. If he sets his cut-off point low, he accepts more signals, but also rejects less false-positives. If he sets his cut-off point high, he accepts less false-positives, but rejects more signals.

Curves are drawn relating the probability of detecting signals (or-

dinate 0-1.00) to the probability of reporting signal presence when only noise was presented (abscissa 0-1.00). Where there is complete overlap of signal-plus-noise (SN) and noise (N) distributions, S will report Yes with equal probability for both SN and N ; that is, if he says Yes 15% (or, say, 80%) of the time, he will say Yes to this proportion of SN presentations and also of N presentations. The probability of saying Yes will be the same for both situations, producing a diagonal with a slope of 1.00. For a higher signal-noise ratio, that is, making the signal more intense, a concave curve running diagonally is drawn. An even greater signal-noise ratio will produce another curve above this, the parameter being such increase. These curves are called R.O.C. curves (receiver operating characteristic) and ideally represent the "best that can be done with the information available" (177, p. 104).

It is explicit in the R.O.C. curves that as true positives increase so do false positives (called "false alarms"). Any curve is for a given signal-noise ratio, and since it runs diagonally, increase in ordinate values is associated with increase in abscissa values. "A zero false alarm rate can be attained only at the sacrifice of zero detection rate. As the allowable false alarm rate increases, the increase in detection is first rapid and then slower, not a linear function of the false alarm rate," states Birdsall (9, p. 397). Indeed, a "wild observer" willing to produce 28% false positives at threshold "would have been found to have a threshold 6.4 db lower than the equally good but conservative (2%) observer" (p. 397) for an energy level cited. Caution raises the threshold.

Forced-choice tasks involve no such criterion but rather "the largest of a set of likelihood ratios (a rela-

tionship between probability densities of detection and false alarms) is chosen and this will be correct unless one of the likelihood ratios due to noise alone exceeded that likelihood ratio due to signal-plus-noise" (9, p. 401).

From the family of R.O.C. curves generated, the conventional psychophysical curves can be derived. Relation of the conventional threshold to false alarm probability suggests that "the threshold is strictly a monotonic function of the *false alarm probability*" (p. 397, italics supplied).

In addition to giving the usual notion of a criterion, which involves choice of a particular signal-noise ratio with report of any input above it, Birdsall presents others: establishing a permissible level of false positives and increasing positive report until that level is reached; reporting in terms of veridical proportions of SN and N; reducing information (Shannon) uncertainty. Tanner and Swets (177) report experiments in which values and costs were announced to *S*, who could get as much as \$2.00 extra in a session. The *Ss* were given all the information necessary to compute optimization criteria and although they were mathematically naive, they behaved according to decision theory, Swets and Tanner reported (176). The *S* acts, according to Tanner (175) *not to optimize information, but to maximize expected value*. Stated otherwise, *S* is maximizing the consequences of his responses in accordance with risks and profits attached, and with past learnings. The response is an operant and should not be confused with stimulus discrimination (cf. 67).

The conversion of Yes-No curves to forced-choice curves, and the reverse, are reported by Tanner and Swets (177) and Swets, Tanner, and

Birdsall (174). "R.O.C. curves fit the Yes-No data, and forced-choice and Yes-No data are consistent with each other" (175, p. 406). This is a reported rationalization of the two main indicators of the subliminal effect. This effect can involve the processing of information to give "high" accuracy curves while high cut-offs for signal detection produce "low" Yes curves, leading to a discrepancy or asynchrony between them; that is, *S* says No and is accurate. These responses are operants governed by their consequences. Where the consequences are unfavorable for Yes, *S* will produce a large subliminal effect.

Regarding the presence of information in null semantic responses, the evidence for which has been presented in the sections on developments in indicator methodology and decision theory, Tanner considers that "the implications of such presence are tremendous." If such information is not lost, *O* is "capable of ordering the 'below threshold' events. This means that an observer receives more information than he is able to transmit when he has at his command only a one-bit channel" (175, p. 404), that is, only the two alternatives of Yes-No.

Is there information in null scores of the accuracy indicator? Here there is typically either congruence between *S*'s response and *E*'s score sheet, or a null congruence between them. If *S* is either accurate or inaccurate, then giving him a second choice when he is wrong (on the first) should produce no further information. Stated otherwise, congruence at second choice should be what one would expect for pure guessing, that is, chance congruence. If in a four-choice situation, *S* is wrong on the first try, and then tries again, since

there are only three possibilities left, null information would be indicated by .33 correctness at this choice. Swets, Tanner, and Birdsall (174) report a proportion of .46 (significant at far beyond the .00001 level). First choices obtained without allowing for second choice, and first choices obtained with such procedures produced almost identical accuracy for first choice (.650 and .651), indicating that information in the first choice was not thereby affected, and leading Tanner to conclude that "first choices are made up of a certain number of correct choices based on supra-threshold outputs, a certain number of incorrect choices based on supra-threshold outputs, plus some pure guesses" (175, p. 404). Further, "it is not unlikely that the third choice will convey information. Thus it is very difficult for an experimenter to determine when enough information has been extracted from forced choices" (174, p. 54). Accordingly, if "one calculates a threshold taking into account the correct second choices, he must arrive at a value far lower than any so far calculated, and at the same time he must realize that his result is still not as low as the true value. . . . Such considerations suggest for all intents and purposes an arbitrarily small threshold, which is the same as saying that information is continuous," (175, p. 404) Tanner concludes.

The concept of the threshold obviously requires reexamination in the light of this research. The threshold has often been considered as a conceptual stimulus magnitude below which response does not occur, and above which it does. Locating such a point has often proved to be elusive, since the switch in response called for by this interpretation occurs at differing stimulus magnitudes for the

same *O*. This has often been interpreted in terms of errors of measurement or procedure, or in terms of fluctuations in the sensitivity of *S*, in stimulus energy, or both. A formulation which gets around this problem is to consider the threshold as a best estimate of this make-break point. The term is also used without such interpretation as a convenient, but seemingly arbitrary point of 50% response. Setting this convenient point at 50% is, however, not entirely arbitrary, since if fluctuations are normally distributed around a mean, this mean can be the threshold, which can have mathematical significance as the inflection point of the integral of the normal curve. This curve is the psychophysical curve of the phi-gamma hypothesis (cf 181). Conceivably, the possibility of such a mathematical referent explains much of the staying power of the term, threshold, in its current usage.

Turning to the decision curves discussed, no comparable mathematical significance can readily be attached to this point, nor is there a conceptual significance of a best estimate where information is considered a continuous process. If we use as an example the anonymous ditty about the sad fate of Psychology, who first lost her soul, then her mind, then consciousness, being left with only her behavior (of which the less said, the better), we might wonder where Limen is headed.

Some Concluding Remarks on Subliminal Perception

A methodological analysis of subliminal perception experiments in terms of psychophysical indicator methodology indicates that subliminal perception experiments have as their paradigm a discrepancy or asynchrony between two indicators,

one being a semantic indicator, and the other varying. The greatest support for the subliminal effect comes from experiments where the second indicator is an accuracy indicator. These two indicators were systematically compared. One provides considerable control by the experimenter, the other does not. Whereas false positives are usually corrected for in standard psychophysical research, false negatives, in a logical inconsistency, are not corrected for; their existence defines the subliminal effect. This procedural inconsistency can of itself produce a subliminal effect.

Recent research by Blackwell and others submitting the two indicators to systematic experimental comparison suggests that semantic indicators tend to admit more variance extraneous to discrimination than do accuracy indicators. They are therefore less sensitive and will produce less response, creating a subliminal effect as an artifact of this contamination. Among invalidating variables discussed were consequences of response, serial effect, and categories of report.

Concern with these two indicators has also been evident in experimentation related to decision theory. This work questions many of the assumptions of a threshold, upon which the notion of subliminal perception rests. Response in a Yes-No situation is viewed as the outcome of a decision process, with the *S* having available to him far more information than he reports. Rather than seeking to maximize information, he optimizes consequences of his response, in line with work on operant conditioning. Decision variables extraneous to sensory discrimination will influence his response. Analysis of null accuracy responses supports the con-

tention that accuracy is not an all-or-nothing phenomenon, and suggests that thresholds are far lower than hitherto considered. There is considerable evidence to suggest that the "below threshold" of subliminal is considerably above threshold. A merchant who alters price-tags to report wholesale price at much higher than it is, can sell "below cost" and make considerable profit.

In sum, methodological analysis suggests, to borrow a phrase from the political arena, that, regarding the subliminal effect, "there is less here than meets the eye."

II. SUBCEPTION, PERCEPTUAL DEFENSE, AND VIGILANCE

The term *subception* first appeared in a report of an asynchrony obtained between two indicators (109), and is similar in this respect to the subliminal perception experiments discussed. Lazarus and McCleary (105) employed visually presented nonsense syllables, some of which were conditioned to shock. These syllables produced GSR's when *S* was reported unable to identify them correctly. Subception was defined as "a process by which some kind of discrimination is made when the subject is unable to make a correct conscious discrimination." It will be noted that awareness was not inferred from a semantic indicator, as it is in subliminal perception, but from an accuracy indicator. This indicator is involved in the vast majority of subception, perceptual defense, and perceptual vigilance studies, and has indicator properties which are singular enough to merit a methodological analysis of its own, as a sub-branch of subliminal perception.

In general, these studies report a threshold stimulus magnitude for cor-

rect identification of a stimulus such as a word or a drawing. This threshold is related to other variables. In perceptual defense and vigilance studies, differences in thresholds are systematically related to different classes (examples are given in parentheses) of stimuli (socially acceptable words, nonacceptable words), subjects (schizophrenics, depressives), states (stress, nonstress), experimental procedures (allowing completion or noncompletion of tasks), or various interactions between these and other variables. In the subception studies, an autonomic indicator may be paired with this accuracy indicator, and an asynchrony is obtained (autonomic responses paired with null accuracy). This is related to variables of the type mentioned.

Currently, there is controversy as to whether the differences obtained relate to perception in the experiential sense (cf. 16), to set (cf. 63), to learning processes (cf. 45), to statistical and other artifacts (cf. 21, 83). The battle lines drawn are between the experiential proponents and all others, with the former, so to speak, playing the field. "What one sees, what one observes, is inevitably what one selects from a near infinitude of potential percepts," opens a key-noting experiential article. Perception is "a first line of defense against would-be catastrophic situations and a sensitizer to adaptive opportunities," the discussion continues (140). The two processes mentioned in this quotation are the categories into which experimental results considered positive are placed. *Defense* is related to perceptual insensitivity inferred from high thresholds, and *sensitization* or *vigilance* is inferred from low thresholds. The psychophysical complexities of the stimulus materials utilized, and differences in

procedures involved make it difficult to compare experiments in terms of differential thresholds, or to obtain norms. Accordingly, the thresholds are not high and low in reference to some standard, but in reference to another stimulus group or subject group run during the experiment. Wherever such a group needs to be specified, it will be placed in parentheses, preceded by *vs.* Thus, the statement, "Defense effects were found for schizophrenics (*vs.* normals)," should be read to mean that schizophrenics obtained higher thresholds for the stimuli of the experiment than did normals.

In this discussion, the terms *defense effect* and *sensitization effect* will be used to refer to such high and low thresholds in their systematic relationship to variables of the experiments. As in the case of *subliminal effect*, these terms are used for convenience only and imply neither acceptance nor rejection of any experiential inferences attached to them. As in the preceding section, this discussion will not be concerned with the justifiability of attaching surplus meaning to concept-perception, this being a philosophic issue outside the scope of this discussion. Rather, the discussion will focus on psychophysical indicator methodology, which is minimally common to all definitions of perception. The schematic definition is minimal since if the indicator is invalidated as an indicator of concept-perception, then it is also invalidated as an indicator of perception defined in any surplus manner. The reverse does not necessarily hold. And, as in test theory, an indicator is invalidated by its admission of extraneous variance, in this case, by its correlation with variables historically considered extraneous to the definition of perception. Studies will be

classified according to indicator and method used, and categorized according to substantive issues.

Results from Studies Using the Ascending Method of Limits and Indicator of Total Initial Accuracy

Use of a specific indicator and specific psychophysical method is so widespread in this area as to make these almost generic for the studies under consideration. The method is the ascending Method of Limits, and the indicator of perception is one of initial total identification. An unknown stimulus configuration is first presented at a magnitude considered too low for identification. Stimulus magnitude is then increased in successive presentation until *S* identifies the configuration in a way considered total, that is, responds with the total word on *E*'s score sheet. The stimulus magnitude at which such identification occurs, usually initially, is called the *recognition threshold*.

The most common method for varying stimulus magnitude is through control of duration using a tachistoscope (3, 10, 23, 30, 39, 40, 50, 51, 63, 64, 65, 73, 75, 86, 95, 96, 98, 101, 112, 114, 116, 117, 119, 126, 130, 134, 138, 139, 140, 141, 147, 150, 151, 153, 163, 164, 165, 167, 168, 179, 180, 184); indeed, a photograph of a tachistoscope appeared in an introductory psychology textbook issued during this period (76, p. 297). Variation of voltage for visual stimuli has been used in other experiments (53, 56, 68, 74, 106, 111, 136, 142, 143, 144, 158, 165, 166, 171, 189), resistance in another (125), episcotister angles have been varied (196). Acoustical equipment has been varied in voltage (15, 185), and the intensity of tape recording has been varied (97, 99). Some novel ascending procedures have been carbon copies of typewritten words, from

the most smudged to most legible (7, 24, 35, 36, 37, 191), anatomical drawings ascending from a single line to a fully drawn banana inserted into a mouth (107), Gottschaldt figures, from most broken to most full (160) and progressive alteration of focus (62, 161). The discussion to follow concerns studies using this indicator-method and discussion related to the issues they raise.

Reaction to two early studies initiated a pattern of controversy and accommodation which has characterized the field ever since. A subception study by McGinnies (112) reported GSR's upon presentation of taboo words (whore, penis) at stimulus magnitudes which produced null total accuracy, and also reported a defense effect for these words (vs. acceptable words). A study by Postman, Bruner, and McGinnies (140) reported sensitization effects for words related to *Ss'* values (vs. words unrelated), that is, *Ss* scoring high in economics on a value scale had low thresholds for economic words, with similar relations for other value groups. The experiential interpretation made in terms of recognition and nonrecognition was challenged by Howes and Solomon (85, 86) who argued that an extraneous variable, word frequency, had been introduced, since the words had not been equated for frequency of usage (164). They also charged suppression of taboo words. McGinnies (113, 117) disagreed with the suppression interpretation and reported obtaining defense effects even when frequency was controlled. Equation for frequency by Postman and Schneider (146) in the values experiment eliminated the sensitization effect for words of high frequency, but not of low frequency.

The following studies concern frequency as a methodological variable. Postman and Conger (142) reported

that for low frequency stimulus words *S* responded with words that were more frequent. Blake and Vanderplas (15) reported more errors before high threshold words (vs. low threshold). Long words (vs. short words) were more affected by frequency, reported McGinnies, Comer, and Lacey (116). The amount of information in a word, rather than its length was related to recognition thresholds by Miller, Bruner and Postman (120). DeLucia and Stagner (44) noted frequency effects, but also threshold dependency on variables affecting homeostasis. Vanderplas (184) similarly found frequency effects but also Gestalt effects of the "organized character of the trace systems involved" (p. 582). Goodstein (73) found that an $r = .53$ between unpleasantness and thresholds dropped to $-.18$ with frequency controlled. Newton (129) found more errors for unpleasant words, even with such control.

Differing frequencies were "built into" an experiment by Solomon and Postman (165). Nonsense syllables were presented at differing frequencies (from 1 to 25) during a training period; accuracy threshold was a function of such frequency, as it was in a similar experiment by King-Ellison and Jenkins (95) who obtained an $r = -.99$ between threshold and log frequency. Postman (134) varied training frequency and structural similarity. Frequency lowered thresholds for words of low similarity, but raised them for high similarity words. Postman and Rosenzweig (144) varied training and recognition modalities (auditory and visual); there was facilitation when the same modalities were used, and for auditory thresholds when training was visual, but not the reverse. Noble (131) reported that *Ss* rated as more familiar words which

had been presented more frequently in a training period. Kristofferson (98) found recognition thresholds inversely related to Noble's *m*, which is contingent on frequency and associations. Frequency has also been related to intelligibility (cf. 84), Howes assigning 69% of intelligibility variance to frequency for words studied.

Currently, the Thorndike-Lorge list is used as a control, as in an experiment by McGinnies and Sherman (117). Postman and Rosenzweig (145) report threshold frequency relations for French words. Davids (41) reported that frequency of association to words related to their personal relevance, not to "frequency per se" (p. 335). Similarly, Daston (40) found sensitization effects for words frequently given by *Ss* in therapy, giving him grounds for support of idiosyncratic rather than "usage table" frequency.

Defense effects have been found in the McGinnies studies cited (112, 113, 117), by McGinnies and Adornetto (114) for both schizophrenics and normals to sexual words (vs. neutral), by Cowen and Beier (35, 36) who used blurred carbon copies of sexual words (vs. neutral), and by Kleinman (97) using emotional words (vs. neutral) for the psychogenic deaf (vs. organic). Eriksen (52) found defense effects by paranoids against hostile words (vs. other categories). Spielberger (168) found thresholds dropped as the experiment progressed except where stutterers (vs. non-stutterers) responded verbally (vs. in writing) for stutter-arousing (vs. non-arousing) words; thresholds continued high here, leading *E* to assign this defense effect to response suppression.

On the other hand, *sensitization* effects have been found to *double-entendre* words (fairly, pussy, balls,

screw) in a sexually connoting list (vs. nonsexual) by Wiener (191), by Kurland (99) for emotional words (vs. neutral) for hysterics and obsessives (vs. normals), by Lindner (107) for sexual pictures (vs. nonsexual) for sex offenders (vs. other criminals), and by Daston (39) for homosexual words (vs. others) for paranoids (vs. schizophrenics), in contrast to Eriksen (52) who did not. Chodorkoff (27) reported sensitization to threatening words (vs. neutral) by better-adjusted Ss in therapy who were "getting to know, as quickly as possible, what it is that is threatening" (p. 511). In another study (29), only absolute change (defense or sensitization) was significant. McClelland and Liberman (111) found neither support nor rejection for hypothesized relations between thresholds and *n* Achievement, but Eriksen (51) reported sensitization to aggressive pictures (vs. nonaggressive) by Ss giving aggressive T.A.T. stories. Greenbaum (74) reported high anxiety Ss (vs. L.A.) displayed sensitization to hostile faces.

The argument has been raised (cf. 78, 87) that these effects may be *ad hoc*, since given two groups to be compared, if they are not equal, one must be higher than the other. Accordingly, for any inequality, a substantive result of defense or sensitization *must* be obtained. The strength of a hypothesis being the ease with which it can be disproven, the need-perception hypothesis may be jeopardized. Spence (166) and Chodorkoff (29) have argued that either outcome is substantive. Where defense and sensitization are not explanatory concepts, Postman (135) states, they are legitimate opposing principles; *E*, however, should "anchor the concept in *antecedent*" con-

ditions (p. 299). Such an experiment is reported by Stein (171) who first classified Ss as sensitizers and defenders, and *then* obtained corresponding threshold effects, as did Carpenter, Wiener, and Carpenter (24). The Ss diagnosed as inhibited displayed defense effects against sexual words, and reported response suppression, Kissin, Gottesfield, and Dicks (96) found. Cowen, Heilizer, Axelrod, and Alexander (37) found no relation between Taylor Anxiety scores and perceptual scores.

The words used visually in the Postman, Bruner, and McGinnies (140) study of values were transcribed verbally for a tape recorder by Vanderplas and Blake (185) with similar sensitization effects. The Ss ranked traits; Haigh and Fiske (75) obtained *defense* effects for liked traits (vs. least liked). Rosenthal (152) noted differences in use of *value* related to religion, science as against likes, dislikes; the Allport-Vernon scale of values initially used had been considered as confounding *interest* and *value* by Adams and Brown (2). Traits named desirable also produced defense effects (vs. undesirable), Postman and Leytham (143) reported. Negative results were obtained by Gilchrist, Lude-man, and Lysak (68) for Ss rated on an Anti-Semitism scale (high vs. low), with nouns projected (Jew vs. Ink), accompanied by adjectives (opprobrious, approving, neutral). *Evidence* of withholding (rather than defense against) derogatory words such as nigger could be elicited "only in unusual circumstances" (p. 733) such as Negro *E* running Negro *S*, Whittaker, Gilchrist, and Fischer (189) reported.

McClelland and Atkinson (110) noted an increase in food responses as a function of deprivation. Wispé

(195) reported this relationship was nonlinear. Sensitization effects to food words (vs. neutral) for deprived Ss (vs. little deprivation) were reported by Wispé and Drambarean (196). Taylor (180) failed to find such a relationship.

Capital letters related to experimentally-induced failure (vs. non-failure) yielded defense effects for Postman and Brown (137). Stressed Ss (vs. nonstress) exhibited defense effects against Gottschaldt figures gradually made more complete by Smock (160); response perseveration was also displayed. Postman and Bruner (138) reported that badgered Ss responded with aggressive and escape words. Eriksen and Browne (56) related sensitization effects for anagrams to learning principles rather than to perception. Both sensitization and defense effects were obtained by Postman and Solomon (147) in a Zeigarnik effect study for completed and uncompleted tasks. Eriksen (53) found that Ss who forgot failure-associated words also exhibited defense effects to them. Spence (167) reported sensitization to failure words.

In the following studies, presentation of a visible word, the standard, is followed by a word to be recognized using the ascending Method of Limits. Neutral recognition words produced defense effects when the standard was taboo (vs. acceptable), McGinnies and Sherman (117) reported, adducing this as evidence for a perceptual rather than response-suppression effect. High anxiety Ss (vs. low) exhibited defense when the standard was threatening (vs. non), Smock (163) reported. Cofer and Shepp (30) varied synonymy of standard and variable, reporting lowered thresholds for more synonymous (vs. less) words. Taylor

(179) obtained defense when Ss were not told of a relation between variable and standard (vs. informed).

In a training session, Rigby and Rigby (151) associated token awards [5, 2, 0] and withdrawals [-3] with words. The threshold values of such words ranked themselves, lowest first, as 5, 2, -3, 0. Newton (130) supplied nickels [1, 0, -1]. There were defense effects against withdrawal words.

Pronunciation of a given word terminated shock; Reece (150) reported sensitization effects (vs. non-termination word).

Neisser (126) gave Ss a list of words which *E* said would appear. There was sensitization to these words rather than to homonyms and unrelated words. Since homonyms involved the same responses as test words, *E* concluded that set "facilitates the perception of specific visual patterns" (p. 402), that is seeing, not saying. Ross, Yarczower, and Williams (153) varied similarity of homonyms (be-bee vs. phrase-frays), obtaining a nonlinear relationship to thresholds.

Defense effects have been obtained when common situations have been rendered uncommon (vs. expected situations). Thus, Bruner and Postman (23) presented playing cards with, for example, red spades. Postman, Bruner, and Walk (141) reversed letters. Smock (161) presented figures half-man, half-woman. The Ss required to discriminate along one dimension had lower thresholds than in situations where stimuli could belong in two categories, Postman and Bruner (139) reported. Similarly, Freeman and Engler (65) set one group for color words, another for food *or* color. Multiple set lowered thresholds for low frequency words, but raised them for high.

Engler and Freeman (50) set one group for animals, another not. An "impairment of recognition" of words not appropriate to the set was reported. Stein (172) reported that as exposure time of Rorschach cards increased, so did $F+$ responses. This procedure enabled study of "the perceiving process . . . as it unfolds itself" (p. 356).

Generally, where S has been given some foreknowledge (vs. no foreknowledge) of words to be shown, defense effects have disappeared: for obscene (vs. acceptable) words used by Lacy, Lewinger and Adamson (101) and similarly for taboo (vs. acceptable) words presented by Postman, Bronson, and Gropper (136), for hostile (vs. neutral) words by Smith (158), for taboo (vs. fruit, neutral) by Freeman (63, 64) who reported "no evidence for perceptual defense against taboo words when S s have been set by instructions to look for and report such words" (63, p. 285). Mausner and Siegel (119) found no relation between monetary value assigned to postage stamps and thresholds, but the session assigning arbitrary values should be considered as constituting a foreknowledge session. Development of insight was the interpretation given by Bitterman and Kniffin (10) to the threshold drop in taboo (vs. neutral) words as the experiment progressed. Chodorkoff (28) asked why insight had not developed toward neutral words. Beier and Cowen (7), however, reported defense effects to emotional (vs. neutral) words (whore, anger) even with forewarned S s, as did Aronfreed, Messick, and Diggory (3), for unpleasant words (vs. pleasant) with forewarning. Lawrence and Coles (103) presented a list *after* stimulus presentation as well as before. Both groups had equally lowered

thresholds (vs. control), leading the E s to conclude that response facilitation and memory were involved.

Studies Using Forced-Choice Indicators

A group of studies reports use of a "psychophysical method, adopted from Blackwell . . . (who has) reported detection thresholds of a highly reliable character obtained with this method" (4, p. 39), that is, spatial forced-choice, with four positions. Complex stimulus configurations are generally presented, with at least one picture relevant or made relevant to S . Sensitization is inferred from increased call of a location containing this critical picture (usually vs. non-critical), and defense from such decreased call.

Blum (16) presented dogs from his Blacky test. The S reported which *stood out best*. A critical dog (masturbating, oral-aggressive) and neutral dog (carefree) were shown with S told to consider "when *you* might have felt" this way (p. 95). The perceptual task was then repeated at a duration representing the same "low level of awareness" (p. 95), giving a sensitization effect for the critical dog (vs. first test) predicted from psychoanalytic theory. Exposure time was then increased to one involving previous recognition of "one or more of the pictures" (p. 96). The S was asked to locate the critical dog and was *scored for accuracy*. Accuracy frequency was lower than preceding standing-out frequency in accord with a predicted repression (defense effect). Smock (162), arguing that in a complex stimulus the "properties which determine whether the stimulus 'stands out' clearly at .03 seconds" are not identical to those "which determine the ease of correct discrimination at .20 seconds" (p. 70), attempted to

provide such a control, and obtained sensitization effects, with analogous defense effects not as clear.

In a second experiment, Blum (17) classified Ss according to conflict areas of the 11 Blacky pictures. The same four dogs were always shown. The S was told all 11 would appear equally and was to *identify each* dog presented "subliminally" (p. 25). The score was not accuracy but number of calls for each of the 11 dogs. Critical dogs not shown were called with frequencies similar to neutral dogs not shown, arguing against response suppression. There was, however, a significant defense effect against critical dogs *projected* (vs. neutrals projected), hence Blum argued "perceptual defense can be traced to the perceptual process itself" (p. 27), which can be considered unconscious since "subjective reports show that the phenomenon was not a conscious one" (p. 27).

An experiment by Nelson (127) combined procedures used in both Blum experiments. Defense effects were obtained with the call-frequency indicator, and sensitization effects with the stand-out indicator. Blum (18) called out a location, with S required to say "Blacky doing so-and-so" (p. 171). Number of calls and accuracy were compared. Where accuracy was low, there were less calls for critical dogs (vs. neutral). Where accuracy was high, there were more calls for critical dogs (vs. neutral), leading E to suggest a "defense against a defense" (p. 173).

Atkinson and Walker (4) presented a face (critical) and lamps, plates, and the like, using the stand-out indicator. High n Affiliation Ss displayed sensitization to faces (vs. low n Affil). Pustell (148) used a similar indicator with geometric figures, the critical having been associated with shock.

Males displayed vigilance effects (vs. neutral figures) and females tended to defense: "one way to reduce the drive was to avoid seeing its stimulus wherever possible" (p. 434).

Dulaney (45) attached differential consequences to responses involving which of four positions stood out best. In one case locating a critical figure led to shock; in the other *not* placing it did so. Geometric figures were presented at a "level of awareness too low to be named." Critical places dropped in the first case and rose in the second, leading E to conclude that "perceptual defense and vigilance are learned reactions to anxiety arousing stimuli" (p. 337).

Goldiamond (71) scored Ss on accuracy of forced-choice location of a triangle started at 0 intensity and gradually increased. The S received immediate information on accuracy, and had to correct himself. An experimental group was told ESP was involved. As the stimulus increased in magnitude, E argued, these Ss would learn that its position was related to reinforcement, and when this occurred, S would start responding to the stimulus. Since S^D would then be at high magnitude, the curve would take a sudden spurt. This sharply inflected insight-type curve was obtained; controls produced psychophysical ogives.

Studies Using Miscellaneous Indicators

Hypothesizing that shock might produce a "startle response (which) . . . interferes with recognition and recall of briefly presented material" (p. 15), Hochberg, Haber and Ryan (79) sounded a buzzer simultaneously with, and also after, presentation of a nonsense syllable, the buzzer being followed by shock. Both conditions significantly lowered correct reports

of critical (vs. neutral) syllables.

Lowenfeld, Rubenfeld and Guthrie (108) required identification of nonsense words presented at levels of 40-60% accurate identification, the critical words having been previously associated with shock. They obtained higher GSR's to critical (vs. noncritical) words-presented. That the GSR followed the veridical critical word rather than critical word responses was considered supporting subception. Rubenfeld, Lowenfeld and Guthrie (154) presented variations on a rectangle, and on a triangle, shock being applied to a rectangular figure. Veridical shock figures misreported were associated with greater GSR's than misreported veridical triangles, leading the authors to conclude that stimulus generalization had occurred in subception.

Threatening and nonthreatening instructions were employed by Moffitt and Stagner (124) who presented a completed geometric figure, and then tachistoscopically presented random modifications. The groups differed significantly, with Ss under stress "clinging to one interpretation of an ambiguous figure while it is changing to another" (p. 355). Zeitlin (197) trained Ss on nonsense syllables, two of which were coupled with an annoying noise for a punishment group, with monetary gains for a reward group, and neither for a control. The punished and rewarded words received more correct calls than the controls when presented tachistoscopically at the same low intensity-durations; this was related by *E* to a response shift rather than perceptual change since these words were also called out more often when blanks were used.

An experiment by Eriksen and Wechsler (57) related the amount of

information in *S*'s response to the number of responses used, and the consistency with which he used them. Absolute judgment was used for 11 squares differing in size, but with a range of only 4 jnd's. The Ss classified as anxious and nonanxious were almost identical in discriminative information (1.871 and 1.877 bits) and very close to the 2.000 bits possible on the basis of 4 ideal discrimination responses. Anxious Ss, however, were lower in response equivocation, that is, more fixed in the use of responses available to them; both groups, the *Es* stated, perceived similarly, but responded differently. In another study, Eriksen (54) associated shock with the middle square. One group identified the square by 1-11, and the other group identified the middle square by 6, all others being No. GSR's were taken for all Ss. *E* reported that the number of available responses was a significant determinant of the subception effect. Increasing the number of responses affected "verbal-response generalization but has no effect upon the generalization of the GSR" (p. 360). The GSR was considered nondiscrete, while verbal responses were discrete. A continuous lever movement was compared with verbal response in another experiment by Eriksen (55). Verbal responses were 1-11 and Yes-No. The lever was moved through a seen and unseen arc. The different responses correlated with each other and with stimulus changes, there being significant differences among the latter correlations. *E* concluded response and perception were experimentally differentiable; differences between responses did not necessarily reflect perceptual differences, but could reflect response errors.

The assumption of all-or-nothing

identification upon which any total identification indicator rests was challenged in the subception area by Bricker and Chapanis (21) and Murdock (125) as well as by studies by Eriksen and others cited. Bricker and Chapanis asked *Ss* to make additional guesses when wrong, until *E* said "Right." The guesses turned out to be nonrandom. Murdock argued that if "all-or-none identification did occur, the incorrect guesses should be randomly distributed among all possible alternatives" (p. 566). Evidence was found in both experiments for rejection of an all-or-nothing characterization of accuracy. Accordingly, regarding perceptual defense, sensitization, and subception, Murdock concluded that the *Es* were "unable to exclude the possibility that *S* obtained some information from stimuli which were wrongly identified" (p. 571). Voor (187) combined analysis of information in null semantic responses with subception effects. Shock was associated with crucial syllables, which were then presented at levels yielding accuracy from 0-50%. Three different indicators were concurrently employed: the autonomic GSR indicator, a naming accuracy indicator, and a semantic indicator (seen, doubtful, blind guess). Voor reported that *S* tended to use noncrucial responses when in doubt, and from comparison of the autonomic-accuracy asynchrony with the autonomic-semantic asynchrony, concluded that the subception effect rested upon information in null semantic responses.

The foregoing review of the literature is not intended as an exhaustive account of experimentation in this area. The attempt, rather, has been to indicate the variety of methods employed, the scope of variables investi-

gated, and major conclusions and inferences drawn. Since the validity of these conclusions and inferences must rest upon the validity of the indicators used, the discussion will turn to indicator methodology.

INDICATOR ANALYSIS: THE ASCENDING METHOD OF LIMITS AND THE INITIAL TOTAL ACCURACY INDICATOR

The psychophysical method which is almost generic for these studies is the ascending Method of Limits, and the indicator which occupies similar prominence is an accuracy indicator involving initial total identification. This combination has great face validity. If *E* wants to find out the minimal energy level at which *S* will identify a stimulus, the most sensible approach would seem to be to start out with the stimulus unrecognized, and then raise the energy level until it is identified. Once identified, memory may enter as a variable in the response. It will be a contention of this discussion that the face validity of this indicator is not convertible into other types of validity, and that one would have to go far in the experimental literature to find a more invalid indicator of concept-perception than this indicator coupled with the method employed. It will be further argued that many of the differing extraneous variables thus far discovered, for example, frequency and foreknowledge, can be subsumed under the more general variables elicited through an examination of indicator methodology. Such an analysis may also serve to uncover substantive material which has been overlooked in the controversy, and may also serve to suggest further research. Indicator analysis will be presented under appropriate headings.

Response Bias, Accuracy, and Congruence

Accuracy has been defined as the congruence between *S*'s responses and *E*'s score sheet. Presumably, signals follow the patterns on *E*'s key, but this is not necessary. Bias has been defined in the sampling sense, namely, that not all responses will have an equal probability of appearing in the response sample. To indicate how these could interact to produce defense and sensitization effects, an example will be presented: *S* answers a multiple choice examination of the objective type, with 5 alternatives; *E*'s key sheet is punched so that alternative B is correct, straight down the line. Given two groups of *S*s, one with a slight tendency to respond with a B, Group BR, and one without such a tendency, Group OR, then on this basis alone, BR should produce B responses earlier than OR, and *therefore be correct earlier*, that is, at a lower question number. If, now, instead of numbering the questions in ascending integers, 1, 2, 3, . . . , *n*, the usual procedure, *E* numbers them in ascending durations, .01, .02, .03, . . . , *n* seconds, and calls the first congruence the *recognition threshold*, BR should also have a lower recognition threshold expressed in time, voltage, carbon number, or whatever procedure is used. Instead of varying groups, we can vary correctness of choice, and will get a lower threshold for C in a 5-alternative question than in a 10-alternative question, since there are biases toward the middle response.

Such response biases can in and of themselves account for the results cited. Lindner (107) reported that sexual offenders showed greater "perceptual sensitization" to sexual pictures than other prisoners. An ascending series of pictures, from a

line on up, was used. The sexual offenders were "more sexually responsive than were the controls to sexually stimulating, nonsexually stimulating, and to ambiguously constructed test items" (p. 373). As a matter of fact, the mean number of sex responses for the sex offenders was 10.2; for the controls, 3.2. For the same reason, one would expect sensitization effects for words frequently used by *S*s in therapy, as Daston (40) reports. In a like manner, Elliott and Wittenberg (49) suggest that the accuracy with which anti-Semites identify Jewish faces is a function of a response bias to use the label, *Jew*, for when there is a minority of Jews in the sample, the *S*s become inaccurate.

The bias may be idiosyncratic, as these examples suggest, or it may be a linguistic one. It is under this simple bias heading, rather than in terms of recency, association, generalization, competing responses, and the like (cf. 165), that the effect of word-frequency upon recognition thresholds (cf. 82, 83) can be subsumed. Accordingly, it would be expected that "common color words such as *black*, *brown*, and *grey*" will produce congruences earlier (rather than be "recognized more quickly") than "some of the more esoteric words, e.g., *indigo* and *azure*" (139, p. 372). Data by Zipf (198) indicate a logarithmic relationship between the frequency of a word and its rank order of use. Recognition thresholds are rank-order data.

Biases may also characterize cultural groups; psychological terms are far more frequently used by the author than geologic terms, and he would expect congruence rank to follow suit. Needs would also bias frequency, for when thinking of a pay raise, and mustering arguments for

one, monetary terms would become frequent. Indeed, an increase in food responses as a function of hunger has been reported (110). Chodorkoff's (27) report of sensitization to threatening words by *Ss* progressing in therapy might mean that they could now verbalize problem areas. Response consistencies should also be expected. The discovery that *Ss* who forget words associated with failure will have heightened thresholds for them (53) translates simply to the statement that words whose frequency is lowered by failure will display this negative response bias in a task where a stimulus is increased in magnitude as well as in a memory task. Failure can also, for certain *Ss*, increase response bias (167).

Results on the effects of stress can be simply explained if it is assumed that the experimental stress utilized led to response stereotypy. If an *S* under stress starts off with the wrong response (as he should if there is a large number of alternatives), he will persevere with this incongruence and have a "high threshold," that is, be congruent later than an *S* not under stress, as the various studies cited would indicate (cf. 50, 137, 160, 163).

Neisser's (126) conclusion that "seeing rather than saying" is involved is based upon an assumption of response similarities in homonyms; the verbal motor mechanisms may be identical, but it is questionable if this is an exclusive definition of the response. Lay and lei have different frequencies of occurrence.

Conditioning is, of course, a classical way of biasing responses, that is, altering frequency of occurrence so that it is not random. This would subsume the studies in which foreknowledge was employed. Here, it will be recalled, results almost uni-

formly indicate sensitization, rather than defense. Instructions to human *Ss* serve the same function as training procedures for other animals. Thus, training *S* to certain responses can alter frequency ratios to eliminate previous bias (cf. 3, 10, 63, 64, 101, 103, 136, 158), or can create such through training, as in the built-in frequency experiments (62, 95, 134, 144, 165).

Wiener (191) utilized the same words (fairy, balls, pussy, screw) in two different contexts assuming thereby that frequency had been controlled. This increased the bias *toward* the words in the sexual context. The words have little in common on a nonsexual basis; categorization into a discrete sexual category should facilitate generalization as the data on semantic generalization suggest (cf. 100).

The organism enters the perception situation with built-in response biases, that is, he has been shaped by preceding conditionings. Certain of these biases are so regular as to enable us to recognize him by them; presumably personality relates to such biases. This statement should not be taken to mean that personality variance in perception is excluded from affecting perception by the fiat of being defined as extraneous. It does mean that before personality variables can be accepted as affecting concept-perception, it must be demonstrated that their effect is not via a response bias extraneous to perception. The same consideration would hold for other variables such as need, learning, hedonic value, and the like.⁷

⁷ For stimuli presented at the same intensity, Zeitlin (197) argues that response shifts may be sufficient to explain the reported perceptual shifts of autism, perceptual defense, and sensitization. His argument can be reformulated in the following manner.

The effect of the interaction of this factor with partial identification of a discriminated stimulus needs little elaboration. A couple of letters discriminated may provide the occasion for a response which has been previously reinforced under similar conditions. If this response has a higher probability than others, and this bias will lead to quick congruence, *S* will display sensitization effects. If *S*'s bias does not agree with the score sheet, there may be defense effects. Similarly, partial recognition may lead to suppression of a response which in the past has been followed by aversive consequences.

Total Identification

This indicator-method involves total identification of the stimulus for the congruence to be scored, that is

E's presentations and *S*'s responses can be entered into a 2×2 table in terms of the interaction of presentations-responses which for each are (a) *Negative*, that is, irrelevant to a need under consideration (words such as school, church, in a food or taboo experiment), and (b) *Positive*, that is, relevant to the need, whether the need is "positive" (love, steak), or "negative" (hate, bitch). If *E* flashes both negatives and positives equally, and *S*'s response biases are similarly equal, there will be a .25 entry in each cell. Now, assume that the need dimension is food words, *S* is food-deprived, and his responses shift to .70 relevant, and .30 nonrelevant. The entry becomes:

		<i>E</i> presents (as before)	
		Non-	Rele-
		rel.	vant
<i>S</i> now responds	Relevant	.70	.35
	Nonrel.	.30	.15

The changes in entries (from .25 each) can be interpreted in terms of response shift. The *S* is as accurate as he was before (.15 true-negative plus .35 true-positive equals previous .25 true-negative plus .25 true-positive), and as inaccurate, but his responses have shifted.

On the other hand, one can, as in the subliminal effect cited previously, ignore the whole

the complete word must be named. This can be considered an all-or-nothing notion of accuracy with a vengeance. Bricker and Clapanis (21), Murdock (125), and Eriksen (55) have indicated that for the perceptual studies under discussion there is information contained in erroneous responses, and Tanner (175) and Swets, Tanner, and Birdsall (174) indicate more generally for even simpler situations of psychophysical measurement that there is information in errors, that the threshold is arbitrarily small, and that information is continuous, that is, not discretely divided into information and no-information.

Properties of the Stimulus

Stimulus-response relationships can be classified into various cate-

table. Concentrating only on certain entries, *E* can obtain the following "perceptual" shifts, explainable by "unconscious processes":

1. False-positives (Need response when Nonneed presentation): *autism*. Autism has not been discussed (cf. McClelland and Atkinson (110) food experiment cited) since such studies generally utilize measures other than the threshold data to which this discussion is restricted.
2. True-positives (Need response when Need presentation): *perceptual sensitization*.
3. False-negatives (Nonneed response when Need presentation): suppression, repression, *perceptual defense*.

Since sensitization and defense are complements, the increase in sensitization in the example given means a decrease in defense.

Liberties taken with Zeitlin's presentation include reversing rows and columns to make the presentation accord with mental test theory and presenting defense as a false-negative and complimentary on the table to sensitization.

This unpublished dissertation was brought to the attention of the author after he had sent a copy of this discussion for criticism to Donald T. Campbell, of Northwestern University, to whom the author is indebted for his prompt loan of his personal copy of the dissertation.

gories. For example, a stimulus which follows a response and increases the probability of that response is called a reinforcing stimulus. The stimulus preceding that response may be a discriminative one. These statements can be rephrased to speak of the reinforcing property of the stimulus, and the discriminative property (93, 157). A further distinction can be made between operant and respondent conditioning. These distinctions are not always clear in the literature cited. When a shock is presented "simultaneously" with a stimulus, subtle differences in timing can produce shock effects contingent upon response, rather than produce classical conditioning. This will affect results markedly.

Except for the GSR's of subception, the response in these experiments (and most psychophysical and other perceptual experiments as well) is almost invariably an operant. This means that the response given to a discriminative stimulus is not evoked by it, but will occur in the presence of the discriminative stimulus if such response (in the presence of the stimulus) has related to it a history of reinforcement. A triangle presented on a screen may evoke a GSR if it has been respondently conditioned to shock, but there is nothing about it to evoke the verbal response Triangle, Yes, 7; these are operants. Response will depend upon training procedures and reinforcing stimuli, as well as discriminative stimuli. Cessation of the response can define anxiety (cf. 58) as well as nondiscrimination. In short, the response will co-vary with all variables that affect operants.

It follows that if the stimulus magnitude of the discriminative stimulus is too low for discrimination, reinforcement of an ensuing response

may lead to "superstitious behavior" (93, pp. 102 ff.), that is, there will be reinforcement of the response occurring in the presence of a stimulus other than the discriminative one put in by *E*. Response will be occasioned by an "irrelevant" stimulus. Accordingly, stimulus magnitude of S^D not only affects the discriminability of the stimulus, S^D , the usual function attached to magnitude, but also the extent to which reinforcement can shape the discriminative response.

These two properties of stimulus magnitude should be kept separate; one involves discrimination or perception, and the other the application of reinforcement, a learning variable. The statement, "If the correctness of the response depends largely upon the characteristics of the stimulus, . . . this might be called perceptual behavior" (106, pp. 316-317) would seem to blur this distinction. In the experiment by Goldiamond cited (71), these two functions were separated. In this experiment, increase of magnitude of S^D increased the relationship between response to S^D and reinforcement, leading *S* to respond according to S^D rather than according to ESP instructions he had been given. Stated otherwise, once *S* had learned that a stimulus was being presented, the discriminative property of that stimulus came into play, producing a switch from a zero congruence to one at high level. The "Ah-ah" phenomenon reported by Miller (121) under similarly deceptive circumstances may relate to this inflection point.

This suggests another invalidating variable which may be attached to the particular combination of indicator-method under discussion. In the usual discrimination experiment, where the functional relation between

response and stimulus magnitude is the subject of investigation, *S* is well-trained in what it is that he must respond to. Discrimination can then be related to other variables, such as ablation, stress, dark-adaptation, and so on. Using the indicator-method combination under discussion, *S* must learn during the experiment what it is he is supposed to respond to. Conceivably, some of the differences obtained in the various thresholds could be related to differences in learning variables rather than perceptual ones.

Extraneous Variance and Concept-Perception.

It follows from this discussion that the validity of the accuracy indicator involved as an indicator of concept-perception is questionable. Since its apparent validity as an indicator of perception, with any surplus meaning attached to the term, rests upon its validity in the minimal sense, its status as a valid indicator of perception with any other historical definition may be challenged.

Indeed, the review of the literature presented can be reread with somewhat more clarity if one substitutes *diminished response* for defense effect and for raised thresholds, and *increased response* for sensitization effect and for lowered threshold.

INDICATOR ANALYSIS: FORCED-CHOICE AND MISCELLANEOUS STUDIES

The "psychophysical method, adopted from Blackwell" for some of the forced-choice studies to be discussed, because it yields "detection thresholds of a highly reliable character" (4, p. 39) deviates in certain major aspects from Blackwell's procedures (cf. 11, 12, 14). These deviations may admit extraneous variance and thereby may affect the reli-

ability as well as the apparent validity of the indicators used.

Blackwell uses forced-choice as an accuracy indicator. The studies in this section obtaining defense effects (4, 16, 17, 45, 127, 148, 162) use forced-choice as a *semantic* indicator. The difference between these two indicators has been discussed in the section on subliminal perception, and these alternative uses of forced-choice may be clarified by using an example from multiple-choice procedures. A question can read: "1. Who discovered America? (a) Columbus, (b) Balboa, (c) Drake, (d) Raleigh." Or, it can read: "2. Who was the greatest explorer? (a) Columbus, (b) Balboa, (c) Drake, (d) Raleigh." Question 1 utilizes forced-choice as an accuracy indicator, since it has a correct answer, and accuracy presumably indicates knowledge of history. Question 2 utilizes forced-choice as a semantic indicator; there is no correct answer. Choices made may relate to opinions, biases, and other variables relating to decisions, such as, possibly, the nationality of the teacher. This would also be the case when *E* asks which picture stands out best, or the related question of which is clearest? *No* picture is clearest, hence, *no designation can be accurate*. The *Ss* are asked to name pictures, but accuracy is not the prime concern of *E*. This differentiation into accuracy and semantic indicators is important for consideration of indicator validity since Blackwell (14) and other investigators (69, 70) have reported that the semantic indicator is especially prone to covariation with variables extraneous to perception, and Tanner (175) and other investigators in decision theory (9, 159, 176, 177) have reported that the semantic indicator automatically includes decision processes governed

by maximization of *response utility* rather than of information.⁸ These studies tend further to question the meaning of the null response of *S* when he gives subjective reports that he does not see the stimulus, leading to the inference by some of the *Es* cited that the phenomena involved were not conscious.

These considerations, and the lack of control for information in inaccurate responses as well as in null semantic responses, would seem to make it unnecessary to go into detail into the specific experiments, some of which are highly ingenious. This discussion is not concerned with issues raised by the apparent predictive power of theoretical positions handled adroitly, but does question the relationship of the data to perception.

⁸ This discussion concerns classification of the forced-choice indicator used in these studies as a semantic indicator rather than as an accuracy indicator. The apparent validity of Blackwell's forced-choice indicator derives from its methodology, rather than from its use of forced-choice, or its classification as an accuracy indicator. As the ascending Method of Limits studies indicate, use of an accuracy indicator is no per se guarantee of validity. Some of the criteria involved are discussed by Birdsall (8) for his methods; "signal known exactly" is one criterion.

The multiple-choice examination question given should be considered only an example, since it would probably be an accuracy indicator readily admitting extraneous variance. The more conventional forced-choice indicator would involve structuring the wrong answers as identical to each other, with the correct answer differing along the dimension of discrimination. A four choice situation might have (a) signal plus noise, (b) same noise, (c) same noise, (d) same noise. The example given has a correct answer and three *different* wrong answers.

Difficulties of this kind stand in the way of ready application of these procedures to mental test theory. There seems little reason to doubt that application can be made of the notion of information in incorrect answers in multiple choice tests, currently scored entirely on an all-or-nothing basis.

PERCEPTUAL DEFENSE REVIEWED

An enthusiastic supporter of early subliminal perception experiments and their relation to subconscious phenomena was William James, who regarded exploration in the latter, that is, the "extra-marginal and outside the primary consciousness" as "the most important step forward that has occurred in psychology since I have been a student of that science" (90, p. 228). Binet's work with hysterics was cited approvingly, and subliminal perception was one type of incursion into consciousness "of which the subject does not guess the source, and which, therefore, take for him the form of unaccountable impulses to act" (p. 229), a theme repeated in the present flurry over subliminal advertising, and the reported motivation of many of the studies reported in this discussion; it was to this theme that Titchener and Pyle (182) addressed themselves when they concluded apropos Dunlap's (46) subliminal Müller-Lyer effect, "that if the subconscious is to be received into experimental psychology at all, it must find some other means of access than these imperceptible shadows" (p. 109). The enthusiasm of James for these subliminal phenomena was matched by his contempt for the psychophysics of his day; the methods were "laborious"; Fechner was considered as having pedantically "tabulated no less than 24,574 separate judgments" (89, p. 23).

James is mentioned because the correlation between his attitudes toward unconscious perception and psychophysical methodology is a highly negative one; this correlation would seem at best to have become a zero one in recent times. The recent reviews which surveyed, in part, the areas discussed, did not refer to cur-

rent advances in psychophysics, nor is this atypical. Psychophysical methodology may seem on the surface to have very little relevance to human behavior when compared to the substantive personality discoveries of the unconscious perception studies discussed, but it would appear from this discussion that an understanding of psychophysical methodology is highly relevant for understanding the personality studies and for evaluating and discovering what their substantive contributions are.

It is one of the aims of this paper, and others to follow, to draw attention to the significance of current psychophysical research as a methodological tool of great power for the investigation of those variables that students of habitual behaviors and those considered abnormal have long been concerned with. The research in indicator methodology by Blackwell may provide techniques to isolate not only the extraneous variance in the perceptual response related to the historical variables discussed, but also the perceptual variance itself. If needs (or drugs) do affect perception, it should show up here. Some of the implications of the work in decision theory appear to be evident. For example, the willingness to take risks as indicated in placement on R.O.C. curves conceivably relates to a past history of reinforcements and aversive consequences attached to venturesome behaviors. Another commonality that this research has with operant conditioning is the emphasis it places on the consequences of the response. Comment could also be amplified on the relatedness of these areas to an economics which constructs utility tables from the two variables of the organism's behavior and the environmental consequences.

It is no accident that perception

has attracted the interest of investigators of personality. The experiments reviewed, and clinical practice itself (as judged by tests currently in use) would seem to support the view that *S* will reveal a great deal about himself which he would otherwise not do, by the way in which he approaches a perceptual task. The suggestion made here is that how he responds to this situation provides a key to understanding a past history of response biases built in, effective reinforcements, aversive situations, and the like. These are classically not perception variables. This would seem to be an interpretation more in accord with the evidence presented than one which implicates perception in these differences. Garner, Hake, and Eriksen (67) have recently reasserted a necessary distinction between perception and its response indicator (or the response from which it is inferred), and have pointed out that more than one measure may be necessary for an operational definition.

In view of the difficulty of concluding that perception in its narrowest sense is involved in these experiments, it becomes difficult to see how interpretations attaching even more meaning can have support. Such differences in interpretation of the response have not been debated in this discussion, but a note of caution might be introduced to the effect that the addition of surplus experiential meaning to the response may become experimentally hazardous, though this does not follow from logical necessity. Accustomed to attaching common-sense meanings to responses such as "I didn't see," *E* may assume that it is the semantic referent of the response (the visual experience) that entitles him to attach indicator properties to it. *In actuality, it is not the*

common sense referent of the response that determines the extent to which indicator properties can be attached to it, but its methodological adequacy and usage in relation to concept-perception.

The use of phenomenal report indicators in psychophysics is predicated on a long history of successful procedures and data related to their use, and not upon their having a semantic referent. The response, "I see" is not accepted as an indicator of vision when made by a blind man, nor for that matter, by many *Es* when *S* has vision worse than 20/20. Thus, the danger of using the indicator for subjective inferences is that such use tends to lead *E* to classify responses on a semantic basis, rather than in methodological terms. For example, *E* may state: "These responses relate to vision because *S* says he sees the stimulus when I show it." The methodological classification is then not made because of the ease and face validity of the experiential one. This might explain, in part, why personality and social psychologists who have been taught since Freud's day, at least, to regard *S*'s explanation of his experiences and behavior with suspicion, are found to be vigorously defending a position which works out to state that if *S* said he didn't perceive it, then he didn't (why should he lie?). In effect, they wind up defending a spurious operational position which they have detested and rightly battled in their own area, namely, that perception is what the perceptual response measures.

It is questionable whether the investigations discussed bear on perception, and it is precisely in this lack of bearing that the importance of the experiments may lie. Their extraneous variance is relatable to language, learning, personality; the perceptual response seems to be a rich

source of such information. Before the extraneous variance can, however, be utilized, it must be differentiated from the perceptual variance. Both constitute response variance. The extraction process would seem to require greater attention to psychophysical methodology than that indicated by the William James correlation.

SUMMARY AND CONCLUSIONS

A review of the literature pertaining to unconscious processes in perception was undertaken since preceding reviews had not taken into account current developments in psychophysical indicator methodology, which, it was felt, might clarify issues in this area, especially since they concern the two types of indicators most widely used.

The term *indicator* was defined as the response element, the dependent variable, in an experiment conducted according to perception methodology. Validity of this response as an indicator of perception was discussed, with variance demonstrated historically as extraneous to perception being considered as tending to invalidate it. No stand was taken as to whether perception would be considered a concept defined by its methodology, a subjective experience, or a sensation implying such an experience, this being considered a philosophical issue irrelevant to the conduct of research which is methodologically adequate. Validity as an indicator of concept-perception was considered crucial to the discussion, since if the indicator is invalid in this minimal sense, it must also be invalidated as an indicator of perception in any broader sense.

The experimental literature in subliminal perception, subception, perceptual defense, and perceptual sensi-

tization was reviewed in terms of indicator methodology. The greatest support for subliminal perception comes from experiments in which there is a discrepancy or asynchrony between an accuracy indicator and a semantic indicator, such that *S* is accurate at times when he reports no awareness of the stimulus. This has been called "discrimination without awareness." The indicator referents of subception are an autonomic indicator and an accuracy indicator, such that autonomic responses occur when *S* has not identified the stimulus correctly. The indicator referent for most perceptual defense and vigilance studies is the accuracy indicator used for subception. Defense refers to systematically raised thresholds obtained with this indicator, and sensitization to systematically lowered thresholds. The threshold comparison is in terms of classes of stimuli, *Ss*, experimental conditions, or combinations of these.

The accuracy and semantic indicators of subliminal perception were described and systematically compared. Differences between the two relate essentially to the fact that in the accuracy indicator, the response is not scored, but rather the congruence of that response with *E*'s score sheet. The semantic indicator utilizes response scores. These differences necessitate different controls for bias. Among these controls is the standard correction for false positives, that is, reporting Yes in the absence of a stimulus. Corresponding control for false negatives is not made, and this logical inconsistency defines the subliminal effect.

Recent research in indicator methodology was considered; this bears directly upon these two indicators. This research tends to indicate that the semantic indicator readily ad-

mits variance extraneous to perception and, accordingly, yields higher thresholds and lower reliability than accuracy indicators. Research in decision theory was also examined for implications for these two indicators. This research challenges certain of the assumptions upon which much standard psychophysical research is based. Among considerations of decision theory is that *S* can increase true positives, that is, detection, only at the price of increasing false positives. Accordingly, the semantic indicator automatically involves a decision process as to which level of risk will be taken. Experimentation tends to support the conclusion that *S* tends to maximize utility of his response, rather than information. The response can be considered an operant governed by its consequences. Regarding accuracy indicators, research here has supported the contention that congruence is not all-or-nothing, and that there is information present in inaccurate responses.

The accuracy indicator used in subception, perceptual defense and sensitization was investigated. This is coupled with the ascending Method of Limits. It was argued that the results obtained using this indicator-method combination could be accounted for on the basis of response bias. The *Ss* whose response biases were in the direction of the entries on *E*'s score sheet would have a response-entry congruence earlier than *Ss* without such biases, and since an ascending stimulus magnitude is coupled to temporal sequence, would appear to have a lower recognition threshold. Another invalidating extraneous variable considered involves *S*'s being required by this method not only to discriminate the stimulus, but also to learn which stimulus it is prior to discriminating it. Since this

indicator as currently used also requires total recognition for a congruence score, this was discussed in light of research demonstrating information in errors.

Studies using forced-choice procedures were also examined. These procedures were considered as involving semantic indicators rather than accuracy indicators and, accordingly, open to the admission of extraneous variables which tend to invalidate them as indicators of perception.

Accordingly, it is questionable whether the studies cited indicate discrimination without awareness, unconscious processes in perception, and the like, or demonstrate discrepancies in and between indicators. These discrepancies can be functions of pairing an apparently valid indicator with one made less sensitive by

admitting invalidating variance, or by using procedures which artificially inflate thresholds and thereby make it appear that processes related to receipt of information are going on at below-threshold levels.

It is concluded that most of the substantive contributions of the experiments reviewed cannot be demonstrated to be related to perceptual variables, and this is probably where their importance lies. The S, in responding to a perceptual situation, tends to respond in terms of the consequences of his response and in relation to other nonperceptual variables which probably characterize his responses in other areas as well. This information he would probably not give under other circumstances, making the perceptual response a fertile one for investigating behaviors needed for assessment.

REFERENCES

1. ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383-405.
2. ADAMS, J. K., & BROWN, D. R. Values, word frequencies, and perception. *Psychol. Rev.*, 1953, **60**, 50-54.
3. ARONFREED, J. M., MESSICK, S. A., & DIGGORY, J. C. Re-examining emotionality and perceptual defense. *J. Pers.*, 1953, **21**, 517-528.
4. ATKINSON, J. W., & WALKER, E. L. The affiliation motive and perceptual sensitivity to faces. *J. abnorm. soc. Psychol.*, 1956, **53**, 38-41.
5. BAKER, L. E. The influence of subliminal stimuli upon verbal behavior. *J. exp. Psychol.*, 1937, **20**, 84-100.
6. BAKER, L. E. The pupillary response conditioned to subliminal auditory stimuli. *Psychol. Monogr.*, 1938, **50**, No. 3 (Whole No. 223).
7. BEIER, E. G., & COWEN, E. L. A further investigation of the influence of "threat-expectancy" on perception. *J. Pers.*, 1953, **22**, 254-257.
8. BINET, A. *On double consciousness: Experimental psychological studies*. Chicago: Open Court, 1896.
9. BIRDSALL, T. G. The theory of signal detectability. In H. Quastler, *Information in Psychology*. Glencoe: Free Press, 1955, 391-402.
10. BITTERMAN, M. E., & KNIFFIN, C. W. Manifest anxiety and perceptual defense. *J. abnorm. soc. Psychol.*, 1953, **48**, 248-252.
11. BLACKWELL, H. R. Studies of psychophysical methods for measuring visual thresholds. *J. opt. Soc. Amer.*, 1952, **42**, 606-616.
12. BLACKWELL, H. R. The influence of data collecting procedures upon psychophysical measurement of two sensory functions. *J. exp. Psychol.*, 1952, **44**, 306-315.
13. BLACKWELL, H. R. Evaluation of the neural quantum theory in vision. *Amer. J. Psychol.*, 1953, **66**, 397-408.
14. BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Engng. Res. Bull.*, Univ. of Michigan, 1953, **36**.
15. BLAKE, R. R., & VANDERPLAS, J. M. The effect of precognition hypotheses on veridical recognition thresholds in auditory perception. *J. Pers.*, 1950, **19**, 95-115.
16. BLUM, G. S. An experimental reunion

- of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **49**, 94-98.
17. BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24-29.
 18. BLUM, G. S. An investigation of perceptual defense in Italy. *Psychol. Rep.*, 1957, **3**, 169-175.
 19. BOARDMAN, W. K. Utilization of word structure in prerecognition responses. *J. Pers.*, 1957, **25**, 672-685.
 20. BRESSLER, J. Illusion in the case of subliminal visual stimulation. *J. gen. Psychol.*, 1931, **5**, 244-251.
 21. BRICKER, P. D., & CHAPANIS, A. Do incorrectly-perceived stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181-188.
 22. BRUNER, J. S., & POSTMAN, L. Perception, cognition, and behavior. *J. Pers.*, 1949, **18**, 14-31.
 23. BRUNER, J. S., & POSTMAN, L. On the perception of incongruity: A paradigm. *J. Pers.*, 1949, **18**, 206-223.
 24. CARPENTER, B., WIENER, M., & CARPENTER, JANETH T. Predictability of perceptual defense behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 380-383.
 25. CASON, H. The conditioned pupillary reaction. *J. exp. Psychol.*, 1922, **5**, 108-146.
 26. CASON, H., & KATCHER, N. An attempt to condition breathing and eyelid responses to a subliminal electric stimulus. *J. exp. Psychol.*, 1933, **16**, 831-842.
 27. CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, **49**, 508-512.
 28. CHODORKOFF, B. A note on Bitterman and Kniffin's "Manifest anxiety and perceptual defense." *J. abnorm. soc. Psychol.*, 1955, **50**, 144.
 29. CHODORKOFF, B. Anxiety, threat, and defensive reactions. *J. gen. Psychol.*, 1956, **54**, 191-196.
 30. COFER, C. N., & SHEPP, B. E. Verbal context and perceptual recognition time. *Percept. mot. Skills*, 1957, **7**, 215-218.
 31. COFER, C. N., & SHEVITZ, R. Word association as a function of word frequency. *Amer. J. Psychol.*, 1952, **65**, 75-79.
 32. COLLIER, R. M. An experimental study of the effects of subliminal stimuli. *Psychol. Monogr.*, 1940, **52**, No. 5 (Whole No. 236).
 33. COOVER, J. E. Formal discipline from the standpoint of experimental psychology. *Psychol. Monogr.*, 1916, **20**, No. 3 (Whole No. 87).
 34. COOVER, J. E. Experiments in psychical research. *Leland Stanford Junior Univer. Publ. Psych. Res. Monogr.*, 1917, **1**, 1-641.
 35. COWEN, E. L., & BEIER, E. G. The influence of "threat-expectancy" on perception. *J. Pers.*, 1950, **19**, 85-94.
 36. COWEN, E. L., & BEIER, E. G. Threat-expectancy, word frequencies, and perceptual prerecognition hypotheses. *J. abnorm. soc. Psychol.*, 1954, **49**, 178-182.
 37. COWEN, E. L., HEILIZER, F., AXELROD, H. S., & ALEXANDER, S. The correlates of manifest anxiety in perceptual reactivity, rigidity, and self concept. *J. consult. Psychol.*, 1957, **21**, 405-411.
 38. CRASILNECK, H., & MCCRANIE, E. J. On the conditioning of the pupillary reflex. *J. Psychol.*, 1956, **42**, 23-27.
 39. DASTON, P. G. Perception of homosexual words in paranoid schizophrenia. *Percept. mot. Skills*, 1956, **6**, 45-55.
 40. DASTON, P. G. Perception of idiosyncratically familiar words. *Percept. mot. Skills*, 1957, **7**, 3-6.
 41. DAVIDS, A. Personality dispositions, word frequency, and word association. *J. Pers.*, 1956, **24**, 328-338.
 42. DAVIS, R. C. Motor responses to auditory stimuli above and below threshold. *J. exp. Psychol.*, 1950, **40**, 107-120.
 43. DE LASKI, E. On perceptive forms below the level of the two-point limen. *Amer. J. Psychol.*, 1916, **27**, 569-571.
 44. DELUCIA, J. J., & STAGNER, R. Emotional vs. frequency factors in word-recognition time and association time. *J. Pers.*, 1954, **22**, 299-309.
 45. DULANEY, D. E. Avoidance learning of perceptual defense and vigilance. *J. abnorm. soc. Psychol.*, 1957, **55**, 333-338.
 46. DUNLAP, K. The effect of imperceptible shadows on the judgment of distance. *Psychol. Rev.*, 1900, **7**, 435-453.
 47. DUNLAP, K. Some peculiarities of fluctuating and inaudible sounds. I. The effect of physical interruptions in subliminal phases. *Psychol. Rev.*, 1904, **11**, 308-318.
 48. DUNLAP, K. & WELLS, G. R. Some experiments with reactions to visual and

- auditory stimuli. *Psychol. Rev.*, 1910, 17, 319-335.
49. ELLIOTT, D. N., & WITTENBERG, B. H. Accuracy of identification of Jewish and non-Jewish photographs. *J. abnorm. soc. Psychol.*, 1955, 51, 339-341.
 50. ENGLER, JEAN, & FREEMAN, J. T. Perceptual behavior as related to factors of associative and drive strength. *J. exp. Psychol.*, 1956, 51, 399-404.
 51. ERIKSEN, C. W. Some implications for TAT interpretation arising from need and perception experiments. *J. Pers.*, 1951, 19, 282-288.
 52. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.*, 1951, 46, 557-564.
 53. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, 47, 230-285.
 54. ERIKSEN, C. W. An experimental analysis of subception. *Amer. J. Psychol.*, 1956, 69, 625-634.
 55. ERIKSEN, C. W. Prediction from and interaction among multiple concurrent discriminative responses. *J. exp. Psychol.*, 1957, 53, 353-359.
 56. ERIKSEN, C. W., & BROWNE, C. T. An experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, 52, 224-230.
 57. ERIKSEN, C. W., & WECHSLER, H. Some effects of experimentally induced anxiety upon discrimination behavior. *J. abnorm. soc. Psychol.*, 1955, 51, 458-463.
 58. ESTES, W. K., & SKINNER, B. F. Some quantitative properties of anxiety. *J. exp. Psychol.*, 1941, 29, 390-400.
 59. FERNALD, GRACE M. The effect of the brightness of background on the appearance of color stimuli in peripheral vision. *Psychol. Rev.*, 1908, 15, 25-43.
 60. FERREE, C. E., & RAND, G. Colored after-image and contrast sensations from stimuli in which no color is sensed. *Psychol. Rev.*, 1912, 19, 195-239.
 61. FERREE, C. E., & RAND, G. Contrast induced by color so far removed into the peripheral field as to be below the threshold of sensation. *J. gen. Psychol.*, 1934, 11, 193-197.
 62. FORREST, D. W. Auditory familiarity as a determinant of visual threshold. *Amer. J. Psychol.*, 1957, 70, 634-636.
 63. FREEMAN, J. T. Set or perceptual defense? *J. exp. Psychol.*, 1954, 48, 283-288.
 64. FREEMAN, J. T. Set versus perceptual defense: A confirmation. *J. abnorm. soc. Psychol.*, 1955, 51, 710-712.
 65. FREEMAN, J. T., & ENGLER, JEAN. Perceptual recognition thresholds as a function of multiple and single set and frequency of usage of the stimulus material. *Percept. mot. Skills*, 1955, 5, 149-154.
 66. FRIEDLINE, C. L. The discrimination of two cutaneous patterns below the two-point limen. *Amer. J. Psychol.*, 1918, 29, 400-419.
 67. GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, 63, 149-159.
 68. GILCHRIST, J. C., LUDEMAN, J. F., & LYSAK, W. Values as determinants of word-recognition thresholds. *J. abnorm. soc. Psychol.*, 1954, 49, 423-426.
 69. GOLDIAMDON, I. The relationship of subliminal perception to forced choice and psychophysical judgments, simultaneously obtained. *Amer. Psychol.*, 1954, 9, 378-379. (Abstract)
 70. GOLDIAMDON, I. Serial effect as a function of type of indicator used. Unpublished doctoral dissertation, Univer. of Chicago, 1955.
 71. GOLDIAMDON, I. The ascending method of limits, response availability, and stimulus discrimination. *Amer. Psychol.*, 1956, 11, 419. (Abstract)
 72. GOLDIAMDON, I. Operant analysis of perceptual behavior. Paper read at APA, New York, Sept., 1957.
 73. GOODSTEIN, L. D. Affective tone and visual recognition thresholds. *J. abnorm. soc. Psychol.*, 1954, 49, 443-444.
 74. GREENBAUM, M. Manifest anxiety and tachistoscopic recognition of facial photographs. *Percept. mot. Skills*, 1956, 6, 245-248.
 75. HAIGH, G. V., & FISKE, D. W. Corroboration of personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1952, 47, 394-398.
 76. HILGARD, E. R. *Introduction to psychology*. New York: Harcourt Brace, 1953.
 77. HILGARD, E. R., MILLER, J., & OHLSON, J. A. Three attempts to secure pupillary conditioning to auditory stimuli near the absolute threshold. *J. exp. Psychol.*, 1941, 29, 89-103.
 78. HOCHBERG, J. E., & GLEITMAN, H. Towards a reformulation of the percep-

- tion-motivation dichotomy. *J. Pers.*, 1949, **18**, 180-191.
79. HOCHBERG, J. E., HABER, S. L., & RYAN, T. A. "Perceptual defense" as an interference phenomenon. *Percept. mot. Skills*, 1955, **5**, 15-17.
 80. HOIJER, H. *Language in culture*. Chicago: Univer. of Chicago, 1954.
 81. HOLLINGWORTH, H. L. *Advertising and selling*. New York: D. Appleton, 1913.
 82. HOWES, D. H. On the interpretation of word frequency as a variable affecting speed of recognition. *J. exp. Psychol.*, 1954, **48**, 106-112.
 83. HOWES, D. H. A statistical theory of the phenomenon of subception. *Psychol. Rev.*, 1954, **61**, 98-110.
 84. HOWES, D. H. On the relation between the intelligibility and frequency of occurrence of English words. *J. acoust. Soc. Amer.*, 1957, **29**, 296-305.
 85. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229-234.
 86. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401-410.
 87. HOWIE, D., Perceptual defense. *Psychol. Rev.*, 1952, **59**, 308-315.
 88. HUDGINS, C. V. Steckle and Renshaw on the conditioned iridic reflex: A discussion. *J. gen. Psychol.*, 1935, **12**, 208-214.
 89. JAMES, W. *The principles of psychology*. Vol. I. New York: Holt, 1890.
 90. JAMES, W. *The varieties of religious experience*. New York: Longmans, Green, 1902.
 91. JASTROW, J. *The subconscious*. Boston: Houghton Mifflin, 1905.
 92. JENKIN, N. Affective processes in perception. *Psychol. Bull.*, 1957, **54**, 100-127.
 93. KELLER, F. S., & SCHONFELD, W. N., *Principles of psychology: A systematic text in the science of behavior*. New York: Appleton-Century-Crofts, 1950.
 94. KING, H. E., LANDIS, C., & ZUBIN, J. Visual subliminal perception where a figure is obscured by the illumination of the ground. *J. exp. Psychol.*, 1944, **34**, 60-69.
 95. KING-ELLISON, P., & JENKINS, J. J. The durational threshold of visual recognition as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 700-703.
 96. KISSIN, B., GOTTESFELD, H., & DICKES, R. Inhibition and tachistoscopic thresholds for sexually charged words. *J. Psychol.*, 1957, **43**, 333-339.
 97. KLEINMAN, M. L. Psychogenic deafness and perceptual defense. *J. abnorm. soc. Psychol.*, 1957, **54**, 335-338.
 98. KRISTOFFERSON, A. B. Word recognition, meaningfulness, and familiarity. *Percept. mot. Skills*, 1957, **7**, 219-220.
 99. KURLAND, S. H. The lack of generality in defense mechanisms as indicated in auditory perception. *J. abnorm. soc. Psychol.*, 1954, **49**, 173-177.
 100. LACEY, J. I., BATEMAN, DOROTHY, E., & VANLEHN, RUTH. Autonomic response specificity: An experimental study. *Psychosom. Med.*, 1953, **15**, 8-21.
 101. LACY, O. W., LEWINGER, N., & ADAMSON, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169-174.
 102. LANDIS, C., & VINACKE, W. E. The discrimination of color and form at levels of illumination below conscious awareness. *Fed. Proc. Amer. Soc. exp. Biol.*, 1942, **1**, (1), Part II, 48.
 103. LAWRENCE, D. H., & COLES, G. R. Accuracy of recognition with alternatives before and after the stimulus. *J. exp. Psychol.*, 1954, **57**, 208-214.
 104. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. O. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1950, **19**, 471-482.
 105. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113-122.
 106. LAZARUS, R. S., YOUSEM, H., & ARENBERG, D. Hunger and perception. *J. Pers.*, 1953, **21**, 312-328.
 107. LINDNER, H. Sexual responsiveness to perceptual tests in a group of sexual offenders. *J. Pers.*, 1953, **21**, 364-374.
 108. LOWENFELD, J., RUBENFELD, S., & GUTHRIE, G. M. Verbal inhibition in subception. *J. gen. Psychol.*, 1956, **54**, 171-176.
 109. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: An interim report. *J. Pers.*, 1949, **18**, 171-179.
 110. MCCLELLAND, D. C., & ATKINSON, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205-222.
 111. MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achieve-

- ment on recognition of need-related words. *J. Pers.*, 1949, **18**, 236-251.
112. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244-251.
 113. MCGINNIES, E. Discussion of Howes' and Solomon's note on "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 235-240.
 114. MCGINNIES, E., & ADORNETTO, J. Perceptual defense in normal and in schizophrenic observers. *J. abnorm. soc. Psychol.*, 1952, **47**, 833-837.
 115. MCGINNIES, E., & BOWLES, W. Personal values as determinants of perceptual fixation. *J. Pers.*, 1949, **18**, 224-235.
 116. MCGINNIES, E., COMER, P. B., & LACEY, O. L. Visual-recognition thresholds as a function of word-length and word frequency. *J. exp. Psychol.*, 1952, **44**, 65-69.
 117. MCGINNIES, E., & SHERMAN, H. Generalization of perceptual defense. *J. abnorm. soc. Psychol.*, 1952, **47**, 81-85.
 118. MANRO, H. M., & WASHBURN, M. F. The effect of imperceptible lines on judgment of distance. *Amer. J. Psychol.*, 1908, **19**, 242.
 119. MAUSNER, B., & SIEGEL, A. The effect of variation in "value" on perceptual thresholds. *J. abnorm. soc. Psychol.*, 1950, **45**, 760-763.
 120. MILLER, G. A., BRUNER, J. S., & POSTMAN, L. Familiarity of letter sequences and tachistoscopic identification. *J. gen. Psychol.*, 1954, **50**, 129-139.
 121. MILLER, J. G. Discrimination without awareness. *Amer. J. Psychol.*, 1939, **52**, 562-578.
 122. MILLER, J. G. The role of motivation in learning without awareness. *Amer. J. Psychol.*, 1940, **53**, 229-239.
 123. MILLER, J. G. The experimental study of unconscious processes. In M. L. Reymert, *Feelings and emotions*. New York: McGraw-Hill, 1950.
 124. MOFFITT, J. W., & STAGNER, R. Perceptual rigidity and closure as functions of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 354-357.
 125. MURDOCK, B. B., JR. Perceptual defense and threshold measurements. *J. Pers.*, 1954, **22**, 565-571.
 126. NEISSER, U. An experimental distinction between perceptual process and verbal response. *J. exp. Psychol.*, 1954, **47**, 399-402.
 127. NELSON, S. E. Psychosexual conflicts and defenses in visual perception. *J. abnorm. soc. Psychol.*, 1955, **51**, 427-433.
 128. NEWHALL, S. M., & DODGE, R. Colored after images from unperceived weak chromatic stimulation. *J. exp. Psychol.*, 1927, **10**, 1-17.
 129. NEWTON, K. R. A note on visual recognition thresholds. *J. abnorm. soc. Psychol.*, 1955, **51**, 709-710.
 130. NEWTON, K. R. Visual recognition thresholds and learning. *Percept. mot. Skills*, 1956, **6**, 81-87.
 131. NOBLE, C. E. The familiarity-frequency relationship. *J. exp. Psychol.*, 1954, **47**, 13-16.
 132. PEIRCE, C. S., & JASTROW, J. On small differences of sensation. *Mem. Nat'l Acad. Sciences*, 1884, **3**, (Pt. 1), 75-83.
 133. PILLAI, R. P. B. K. A study of the threshold in relation to the investigations on subliminal impressions and allied phenomena. *Brit. J. educ. Psychol.*, 1939, **9**, 97-98. (Abstract)
 134. POSTMAN, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation: A Symposium*. Lincoln: Univer. of Nebraska, 1953, 59-108.
 135. POSTMAN, L. On the problem of perceptual defense. *Psychol. Rev.*, 1953, **60**, 298-306.
 136. POSTMAN, L., BRONSON, W. C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215-224.
 137. POSTMAN, L., & BROWN, D. R. The perceptual consequences of success and failure. *J. abnorm. soc. Psychol.*, 1952, **47**, 213-221.
 138. POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314-323.
 139. POSTMAN, L., & BRUNER, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369-377.
 140. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142-154.
 141. POSTMAN, L., BRUNER, J. S., & WALK, R. D. The perception of error. *Brit. J. Psychol.*, 1951, **42**, 1-10.
 142. POSTMAN, L., & CONGER, BEVERLY. Verbal habits and the visual recognition of words. *Science*, 1954, **119**, 671-673.
 143. POSTMAN, L., & LEYTHAM, G. Percep-

- tual selectivity and ambivalence of stimuli. *J. Pers.*, 1951, 19, 390-405.
144. POSTMAN, L., & ROSENZWEIG, M. R. Practice and transfer in visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.*, 1956, 69, 209-226.
 145. POSTMAN, L., & ROSENZWEIG, M. R. Perceptual recognition of words. *J. Speech & Hearing Disorders*, 1957, 22, 245-253.
 146. POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, 58, 271-284.
 147. POSTMAN, L., & SOLOMON, R. L. Perceptual sensitivity to completed and incompleting tasks. *J. Pers.*, 1950, 18, 347-357.
 148. PUSTELL, T. E. The experimental induction of perceptual vigilance and defense. *J. Pers.*, 1957, 25, 425-438.
 149. QUASTLER, H. *Information theory in psychology*. Glencoe: Free Press, 1955.
 150. REECE, M. M. The effect of shock on recognition thresholds. *J. abnorm. soc. Psychol.*, 1954, 49, 165-172.
 151. RIGBY, W. K., & RIGBY, MARILYN K. Reinforcement and frequency as factors in tachistoscopic thresholds. *Percept. mot. Skills*, 1956, 6, 29-35.
 152. ROSENTHAL, D. The selection of stimulus words for value: duration threshold experiments. *J. abnorm. soc. Psychol.*, 1955, 50, 403-404.
 153. ROSS, S., YARCZOWER, M., & WILLIAMS, G. M. Recognitive thresholds for words as a function of set and similarity. *Amer. J. Psychol.*, 1956, 69, 82-86.
 154. RUBENFELD, S., LOWENFELD, J., & GUTHRIE, G. M. Stimulus generalization in subception. *J. gen. Psychol.*, 1956, 54, 117-182.
 155. SIDIS, B. *The psychology of suggestion*. New York: D. Appleton, 1898.
 156. SILVERMAN, A. & BAKER, L. E. An attempt to condition various responses to subliminal electrical stimulation. *J. exp. Psychol.*, 1935, 18, 246-254.
 157. SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
 158. SMITH, J. G. Influence of failure, expressed hostility, and stimulus characteristics upon verbal learning and recognition. *J. Pers.*, 1954, 22, 475-493.
 159. SMITH, M., & WILSON, EDNA A. A model of the auditory threshold and its application to the problem of the multiple observer. *Psychol. Monogr.*, 1953, 67, No. 9 (Whole No. 359).
 160. SMOCK, C. D. The influence of psychological stress on the "Intolerance of Ambiguity." *J. abnorm. soc. Psychol.*, 1955, 50, 177-182.
 161. SMOCK, C. D. The influence of stress on the perception of incongruity. *J. abnorm. soc. Psychol.*, 1955, 50, 354-356.
 162. SMOCK, C. D. Replication and comments: "An experimental reunion of psychoanalytic theory with perceptual vigilance and defense." *J. abnorm. soc. Psychol.*, 1956, 53, 68-73.
 163. SMOCK, C. D. The relationship between test anxiety, "threat-expectancy," and recognition thresholds for words. *J. Pers.*, 1956, 25, 191-201.
 164. SOLOMON, R. L., & HOWES, D. S. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, 58, 256-270.
 165. SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, 43, 195-201.
 166. SPENCE, D. P. A new look at vigilance and defense. *J. abnorm. soc. Psychol.*, 1957, 54, 103-108.
 167. SPENCE, D. P. Success, failure, and recognition threshold. *J. Pers.*, 1957, 25, 712-720.
 168. SPIELBERGER, C. D. The effects of stuttering behavior and response set on recognition thresholds. *J. Pers.*, 1956, 25, 33-45.
 169. STECKLE, L. C. Two additional attempts to condition the pupillary reflex. *J. gen. Psychol.*, 1936, 15, 369-377.
 170. STECKLE, L. C., & RENSHAW, S. An investigation of the conditioned iridic reflex. *J. gen. Psychol.*, 1934, 11, 3-23.
 171. STEIN, K. B. Perceptual defense and perceptual sensitization under neutral and involved conditions. *J. Pers.*, 1953, 21, 467-478.
 172. STEIN, M. I. Personality factors involved in the temporal development of Rorschach responses. *J. Pers. Res. Exch. & J. Proj. Techn.*, 1949, 13, 355-414.
 173. STROH, M., SHAW, A. M., & WASHBURN, M. F. A study in guessing. *Amer. J. Psychol.*, 1908, 19, 243-245.
 174. SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. The evidence for a decision-making theory of visual detection. *Tech. Rep. No. 40, Electronic*

- Defense Group*. Ann Arbor: Univer. of Michigan, 1955.
175. TANNER, W. P., JR. On the design of psychophysical experiments. In H. Quastler *Information theory in psychology*. Glencoe: Free Press, 1955, 403-414.
 176. TANNER, W. P., JR. (Chairman) Symposium on signal detection, decision theory, and psychophysical methods. AAAS Convention, Indianapolis, 1957.
 177. TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401-409.
 178. TAYLOR, F. W. R. The discrimination of subliminal visual stimuli. *Canad. J. Psychol.*, 1953, **7**, 12-20.
 179. TAYLOR, JANET A. Effect of set for associated words on duration threshold. *Percept. mot. Skills*, 1956, **6**, 131-134.
 180. TAYLOR, JANET A. Physiological need, set, and visual duration threshold. *J. abnorm. soc. Psychol.*, 1956, **52**, 96-99.
 181. THURSTONE, L. L. The phi gamma hypothesis. *J. exp. Psychol.*, 1928, **11**, 293-305.
 182. TITCHENER, E. B., & PYLE, W. H. The effect of imperceptible shadows on the judgment of distance. *Proc. Amer. Phil. Soc.*, 1907, **46**, 94-109.
 183. TITCHENER, E. B., & PYLE, W. H. On the after-images of subliminally colored stimuli. *Proc. Amer. Phil. Soc.*, 1908, **47**, 366-384.
 184. VANDERPLAS, J. A. Frequency of experience versus organization as determinants of visual thresholds. *Amer. J. Psychol.*, 1953, **66**, 574-583.
 185. VANDERPLAS, J. A., & BLAKE, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18**, 252-266.
 186. VINACKE, W. E. The discrimination of color and form at levels of illumination below conscious awareness. *Archives of Psychol.*, 1942, **38**, (267), 1-53.
 187. VOOR, J. H. Subliminal perception and subception. *J. Psychol.*, 1956, **41**, 437-458.
 188. WEDELL, C. H., TAYLOR, F. V., & SKOLNICK, A. An attempt to condition the pupillary response. *J. exp. Psychol.*, 1940, **27**, 517-531.
 189. WHITTAKER, E. M., GILCHRIST, J. C., & FISCHER, J. W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732-733.
 190. WHORF, B. L. *Language, thought, and reality: Selected Writings*. Cambridge, Mass.: Technology Press, 1956.
 191. WIENER, M. Word frequency or motivation in perceptual defense. *J. abnorm. soc. Psychol.*, 1955, **51**, 214-218.
 192. WILCOTT, R. C. A search for subthreshold conditioning at four different auditory frequencies. *J. exp. Psychol.*, 1953, **46**, 271-277.
 193. WILCOTT, R. C. Subliminal stimulation vs. psychophysical thresholds. *Percept. mot. Skills*, 1957, **7**, 29-36.
 194. WILLIAMS, A. C., JR. Perception of subliminal visual stimuli. *J. Psychol.*, 1938, **6**, 187-199.
 195. WISPE, L. G. Physiological need, verbal frequency, and word association. *J. abnorm. soc. Psychol.*, 1954, **49**, 229-234.
 196. WISPE, L. G., & DRAMBAREAN, N. C. Physiological need, word frequency, and visual duration thresholds. *J. exp. Psychol.*, 1953, **46**, 25-31.
 197. ZEITLIN, L. R. A response oriented analysis of the concepts of autism and perceptual sensitization. Unpublished doctoral dissertation, Northwestern Univer., 1954.
 198. ZIPP, G. K. *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley, 1949.

Received February 10, 1958.

SPONTANEOUS ALTERNATION BEHAVIOR

WILLIAM N. DEMBER AND HARRY FOWLER

Yale University

Beginning with the observations of Tolman (37) and Dennis (6), there has developed over the past three decades a fairly extensive literature on *spontaneous alternation*. Attention to this phenomenon has increased recently with the renewal of interest in exploratory and curiosity behavior. Concern with alternation has had varied bases. To some it represents in simple form a special case of exploratory behavior (e.g., 3, 23). For others it provides a convenient approach to the determination of the course of performance decrement, and in this connection a means of investigating the action of a reinforcer (e.g., 38, 47). A third interest lies in the potential use of alternation as an indicator response for the study of general psychological processes such as memory and perception (e.g., 4, 7, 26). Finally, alternation has been considered a problem area in its own right, worthy of its own theoretical structure (e.g., 14, 38).

Our purpose in this paper is to review the alternation literature in an attempt to codify empirical findings and relate them to theoretical issues. To accomplish this aim, the material has been organized in ways which cut across the areas of interest mentioned above. The contents of the paper are divided into five main sections. Discussion is directed first at some general problems in the definition and measurement of alternation behavior. Then, the major theoretical issues concerning the general explanation of alternation are examined. In the

third section the more specific problem of the sources of alternation, i.e., what is alternated, is considered. The fourth section deals with the data available on variables which affect alternation, and the final section treats of the relation between alternation behavior and learning theory.

It should be noted that this review is confined to the literature on alternation behavior in rats. There are several studies which reveal analogous behavior in human Ss (e.g., 18, 33, 36, 44). But so far these experiments have added little to the information obtained from rat investigations. Alternation in infrahuman species other than rats has been investigated (e.g., 22, 42), but not sufficiently so to warrant review.

DEFINITION AND MEASUREMENT

The typical paper on alternation provides a denotative definition of the sort: "... if an animal turns left in a T maze on its first trial, and if it is immediately returned to the starting point, the probability is quite high that it will turn right on the second trial" (39, p. 19).

More abstractly, alternation implies the occurrence of at least two mutually exclusive categories of behavior over at least two successive time periods. Typically, a "time period" is a trial, and a "category of behavior" a choice-point response in gross, qualitative form. Now, if there are two such categories of behavior, A and B, and two trials are considered, four behavior patterns are pos-

sible: A followed by A, or B by B (repetitions), and A followed by B, or B by A (alternations). If there are more than two possible categories of behavior in a situation, as for example in a maze with three arms, alternation then becomes any non-repetition on successive trials. Whatever the number of trials given, alternation or repetition usually refers to the behavior pattern over any two successive trials.

If alternation is considered purely a response process, the problem of definition is reduced merely to the specification of mutually exclusive responses, e.g., turning right or turning left. But, alternation may involve a reaction to intra- and/or extramaze stimuli instead of, or in addition to, responses. If so, the specification of mutually exclusive categories of behavior becomes somewhat complex. In some situations, alternation of responses may require repetition of stimuli. Conversely, alternation of stimuli may require repetition of responses. Thus, it is often necessary to specify alternation *with respect to* particular stimulus or response dimensions. Just what these dimensions, or sources of alternation may be, of course, is a problem for experimental analysis.

A further problem arises in the statistical treatment of alternation data. Where there are two categories of behavior and two trials, two repetition patterns and two alternation patterns are possible. Under the null hypothesis that the behavior on Trial 2 is independent of that on Trial 1, the a priori probability of alternation, p_o , is .50. The significance of an obtained percentage of alternation would ordinarily be evaluated against this expected value, $p_o = .50$.

Such a procedure, however, involves an assumption which is not always met empirically—namely, that the two categories of behavior are equally likely. The alternation tendency may not be the only factor leading to the nonindependence of Trial 1 and Trial 2 behaviors. There may also exist a repetition tendency. Each *S* behaves in a particular way on Trial 1, and if nothing is changed, should behave in the same way on Trial 2. When, as often is the case, there is initially a strong preference for A or B, $p_o = .50$ overestimates the a priori expectation of alternation. A more reasonable estimate of chance alternation would take into account the distribution of Trial 1 behaviors. Thus, if 80% of the behaviors on Trial 1 were of Type A and 20% of type B, the a priori probability of alternation, p_o , would be $1 - (.80^2 + .20^2)$, or .32.

While rarely found in the experimental literature, this latter statistical procedure is adopted in a recent article by Sutherland (35). An alternative to these statistical manipulations would involve careful control of the experimental conditions so as to minimize unequal distributions of Trial 1 behaviors. Where strong preferences do occur, however, it would facilitate evaluation and comparison of data from different sources if these preferences were used in the estimate of p_o .

GENERAL EXPLANATIONS OF ALTERNATION

In this section our concern is with the most general theoretical problem, i.e., why does *S* alternate? Examination and comparison of the main points of view on this issue will enable a more meaningful evaluation

of the experimental data.

Hull's reactive inhibition. Alternation has been conceived of as purely a response process, explainable via Hull's concept of *reactive inhibition*, or I_R (17). Thus, if a left-turning response is made in a two-choice situation, a certain amount of left-turning inhibition is generated which renders the right-turning response temporarily predominant (47). Integration of the theoretical properties ascribed to I_R leads to deductions of (a) the simple occurrence of alternation behavior, (b) the spontaneous dissipation of alternation over time, (c) a direct relation between alternation and the number of forced trials to one alternative, (d) an inverse relation between alternation and the number of successively massed trials, (e) the response generalization of alternation, and (f) a direct relation between alternation and effortfulness of the response.

Glanzer's stimulus satiation. As was noted previously, conceptions of alternation behavior need not be limited to response-dependent processes. The equally plausible conception of alternation as a stimulus process is exemplified in Glanzer's (14) explanation of the phenomenon. Basic to his system is the concept of *stimulus satiation*, which in formal properties is much like I_R , except that its source is in stimuli, not responses. To quote Glanzer: "Each moment an organism perceives a stimulus-object or stimulus-objects, A, there develops a quantity of stimulus satiation to A" (14, p. 259). After postulating several quantitative relationships, analogous to those ascribed to I_R , Glanzer then states that stimulus satiation reduces the organism's tendency to make any response to A. In this sense, a convenient and repre-

sentative label for the concept is sI .

The similarity in formal properties ascribed to both the I_R and sI concepts is evident in the fact that deductions (a) through (d) of the former system apply also to the latter. The two systems are differentiated, however, by I_R deductions (e) and (f) and by the additional sI deductions which follow: (g) the stimulus generalization of alternation, (h) response repetition, rather than response alternation, with reversal of the stimuli between trials, and (i) a direct relation between amount of alternation and intratrial interval, or length of exposure to an alternative.

Alternation as a manifestation of exploratory behavior. Not very different from Glanzer's view is Montgomery's (23, 24) explanation which makes of alternation a special case of exploratory behavior. Exploration, in turn, is explained as emerging from a *curiosity drive* aroused by *novel stimuli*.¹ The S alternates because the alternative last entered is the less novel.

It is difficult to state precisely how the sI and curiosity drive explanations differ from each other. As a consequence of this ambiguity, or perhaps as another way of emphasizing the ambiguity, there is little research which is directed at narrowing down these two "alternative" explanations. Possibly, the lack of deductive differentiation between the systems is due to an underlying identity which is beclouded by semantic differences. Or perhaps it is due to the failure to consider predictions from the two systems which pertain to aspects of choice-point behav-

¹ The similarity of their concepts to Pavlov's (28) *investigatory reflex* seems not have been explicitly noted by these authors.

ior other than the choice itself.

For example, when an animal is placed in a simple two-choice, non-reinforcement situation, why does it leave the start-box? In accordance with the exploratory drive hypothesis, the novel stimuli beyond the start-box area elicit a curiosity drive which is expressed in the animal's *approaching* and investigating these stimuli. But accumulated experience, e.g., through massed trials, with these stimuli should render their novelty negligible. As a consequence, the animal's latency of selecting an alternative would be expected to increase over trials.

This prediction seems opposed to one deduced from the sI theory. Presumably, in accordance with this view, the animal satiates to the stimuli of the start-box area. Since satiation for certain stimuli reduces the organism's tendency to make any response to those stimuli, the animal *avoids* the start-box area by traversing the stem and selecting one of the alternatives. Over successive trials, the start-box area is experienced every time, but each alternative is experienced only about half the time. As a consequence, satiation for the start-box stimuli exceeds satiation for the alternatives at an increasing rate over successively massed trials. Hence, as trials progress, the increasing tendency for the animal to avoid the start-box area should be reflected in a decrease in the animal's latency of selecting an alternative.

Walker's action decrement. Thus far, our concern with the general explanation of alternation has centered in the distinction between stimulus- and response-oriented theories. However, psychological actions or events need not bear an exclusive relation-

ship to either peripheral response or stimulus events. With this view, Walker (38) proposes, as an alternative to sI and I_R , the concept of an *action decrement*, which is a central event.

Deductive differentiation between the Glanzer and Walker systems appears to be absent except in the extension of the latter model to include roles for levels of motivation and reward. It is hypothesized by Walker that these two factors render greater both the immediate action decrement and a later action increment. Empirically this would mean a greater alternation tendency initially under reward and high levels of motivation, but just the reverse after an extended time lapse.

Other explanations of alternation. The foregoing constitute the major explanations of alternation behavior. However, scattered throughout the literature are others. Alternation has been considered, for example, as an instrumental mode of activity (1, 12), and as a reaction to frustration or punishment (43). These explanations shall not be discussed further since they are unconvincing in the light of the wide variety of conditions yielding alternation behavior. Moreover, so far they have generated little, if any, research.

SOURCES OF ALTERNATION

In the preceding section, we discussed the general explanation of alternation—why the animal alternates—and only indirectly the possible sources of alternation, or what is alternated. Source and motivation are obviously closely related, but nevertheless are distinct issues. Our concern now is with the former.

While the I_R concept provided an

adequate explanation of the typically observed variety of alternation, it could not account for some of the earliest data. In an experiment by Dennis and Sollenberger (9) in which rats were allowed to explore a **Y** maze, the *Ss*' behavior revealed a strong tendency to alternate maze alleys rather than turning responses. When allowed to explore a multiple **Y** maze, in a second experiment, the rats also showed a marked tendency to avoid a pathway occupied shortly before (9). Results similar to these have since been reported by Montgomery (24).

That it was alleys, or more generally stimuli, and not responses that were being alternated in the two-trial situation was an hypothesis tested by Montgomery (25) and by Glanzer (13) in essentially the same way. In both experiments a **+** maze was employed with one leg blocked off on each trial so as to yield a conventional **T** maze. With one leg serving as the starting stem on the first trial, the animal was run from the opposite leg on Trial 2. In this manner stimulus and response components, ordinarily covariate, were separated and pitted against each other. If the *S* alternated responses, it would necessarily repeat stimuli; to alternate stimuli, the *S* would have to repeat its previous turning response. In both experiments the major source of alternation was found to be in the external stimuli rather than in the rat's turning response, in confirmation of the earlier observations of Dennis and Sollenberger (9) and Montgomery (24).

In the Montgomery **+** maze study, the *S*'s first trial response was rewarded. It is therefore conceivable that the second trial behavior ac-

tually manifested response repetition rather than stimulus alternation. Glanzer employed no reinforcement in his experiment, however, thus making the above interpretation unconvincing.

In a subsequent study, which also lacked the use of a reinforcer, Walker, et al. (39) replicated the above results and in addition showed that extra-maze stimuli were less important as sources of alternation than were intramaze stimuli. They proposed that the relative importance of any of the potential sources of alternation, including response, could be varied by manipulation of the discriminability of the alternatives with respect to these sources. Thus, even the response dimension might be increased in importance if the two responses were made more distinct from each other than they are in the typical **T** maze. This hypothesis was tested and confirmed (41). It was further suggested that "response" be thought of as "response-feedback," making it, in this way, another stimulus dimension, rather than something unique. From this point of view, I_R deduction (*e*) becomes a special case of sI deduction (*g*).

Several other experiments have been concerned with determining the sources of alternation. Rothkopf and Zeaman (31) conducted a series of studies which yielded the general conclusion that both responses and stimuli may serve as sources of alternation. They offered the parallel concepts of "tired stimuli" and "tired responses" to account for alternation. In another set of experiments Estes and Schoeffler (12) found little evidence of response alternation and strong evidence of stimulus alternation.

Also pertinent here is the recent finding by Sutherland (35) that the alternation tendency is less when the two goal arms lead to a common goal box than when they lead to separate and distinct goal boxes.² Sutherland argues that alternation occurs not just with respect to choice-point stimuli, but also, and perhaps even more so, it occurs with respect to "the rat's expectancy of the stimuli it will receive beyond the choice point . . ." (35, p. 361).

Overall, the empirical picture shows stimuli as the general source of alternation. The more proximal the stimulus, e.g., intra- as opposed to extramaze stimuli, the greater is its potential as a source. The response may serve as a source of alternation only if its "afferent feedback," a stimulus component, is made salient. In relation to theoretical issues, then, the data support those views which consider a stimulus process as underlying alternation behavior.

VARIABLES AFFECTING ALTERNATION BEHAVIOR

In this section we are concerned with the effect on alternation behavior of a rather heterogeneous set of variables, some of which relate to characteristics of the alternatives and others of which pertain to the temporal duration of alternation.

Amount of work. A crucial deduction, (*f*), of the I_R system led to experiments which investigated the effect on alternation of variations in the amount of work associated with the responses. In all of these experiments, amount of work was not var-

ied between alternatives, but rather over groups of *Ss*, or over different stages of testing for the same *Ss*.

In general, the results of these investigations do not support the I_R explanation.

Mowrer and Jones (27) gave rats access to two equivalent bars in a Skinner box. They found no effect on the rats' alternation between bars of variations in the amount of work required to depress the bars.

Solomon (34) strapped weights on his rats, but found no effect of this manipulation on amount of alternation in a **T** maze. In a second experiment, however, Solomon (34) varied work by inclining the goal arms of a **T** maze 16° from the horizontal, and obtained a slight increase in the amount of alternation. In an attempt to replicate Solomon's positive results, Walker, et al. (41) also used inclined alleys; the rats had to climb a 45° slope in making either response. When compared with rats running in the conventional flat **T** maze, these animals did not show more alternation.

In another **T** maze study, Montgomery (23) failed to find a work effect, but, as he himself points out, the work manipulation came at the end of each alley rather than at the choice point. That is, the amount of work in making the turning response was not varied, and it is not clear why an effect on alternation should be expected under these conditions.

Riley and Shapiro (29) also failed to find any but a slight influence of work on alternation. In their experiment, the work manipulation was effected by varying the weight of doors through which the rats had to push in order to enter a reward chamber.

From a point of view which places

² Dennis (7), however, using a procedure similar to Sutherland's, obtained 80% alternation, which is about the level found in the typical distinct goal-box experiments.

no special importance in the response per se there is no reason to expect work manipulations of these sorts to affect alternation, as long as these manipulations do not mask other differences between the alternatives. On the other hand, if *S* is so burdened down with work that other sources of differential stimulus input are trivial by comparison, then the alternation tendency might be expected to decrease. There is a suggestion of such an effect in the Walker, et al. (41) experiment.

It is possible to apply this interpretation to the results of an interesting experiment by Jackson (19). Rats were run under two conditions: in the first, a conventional open **Y** maze was used; in the second, the same elements of the **Y** were employed except that a 15 cm. gap separated the goal arms from the starting alley. The gap was sufficiently large so that a jumping response was required to cross it. While the usual alternation behavior was observed in the first condition, almost complete repetition was evidenced in the second.

The suggestion above that too much work would lead to a decrease in alternation would apply here if it is further assumed that the rats' jumping was characterized by a strong position bias. The fact that *Ss* were rewarded should increase any such initial repetition tendency. While this explanation seems plausible, replication and further investigation of Jackson's results might best precede continued speculation.

For the sake of completeness we shall note one final type of work-related experiment. Some investigators (12, 46, 47) have attempted to manipulate work by varying the number of forced trials to one alternative. The general result is the expected di-

rect relation between amount of alternation and number of forced trials. While these data have been considered to support the I_R interpretation, they obviously follow equally well from the satiation or novelty points of view. Thus, in general, investigations of the work variable are either not theoretically crucial, or show no effect of work on alternation, or even suggest a relation which is contrary to the I_R prediction.

Pre-exposure. If, as is specified in the sI interpretation of alternation, perception of a stimulus reduces the organism's tendency to respond to that stimulus, then exposure to a stimulus prior to choice should lead to the avoidance of that stimulus if it is soon afterward encountered at a choice point.

Experiments by Glanzer (15) and Sutherland (35) have demonstrated such an effect of pre-exposure. In the Glanzer experiment the *Ss* were placed in one goal arm of a **T** maze, detained there for one minute and then given a choice trial; the *Ss* were either returned to the starting stem by the *E* as is typically done, or were allowed to return there on their own. In both conditions there was a significant tendency for the animals to enter the arm to which they had not been pre-exposed. In Sutherland's experiment, rats were placed and fed in one goal box, and then were given a choice trial. They, too, showed a preference for the alternate arm.

However, with procedures similar to those above, Walker, et al. (40) failed to find any effect. In one experiment rats were exposed to black or white stimulation in small boxes, and then introduced into a **T** maze with a black and a white goal arm. No effect of the prior exposure was evident in the animals' choices. In

another experiment the pre-exposure took place in one of the goal boxes itself. Again, there was no apparent effect on the rats' subsequent choice of goal arms.

It is difficult to account for the discrepancy between the results of the studies cited above. One difference in the procedures followed by Walker, et al. and Glanzer may be relevant. Glanzer's rats were confined in the goal arm and apparently were free to wander within it as far as the choice point; in contrast, the Ss of the Walker, et al. experiment were confined in the goal box and could not return to the choice-point area. Pertinent to this procedural difference is the finding of Kivy, Earl, and Walker (20) that exposure at the *choice point* does influence an animal's subsequent performance.³

In the Kivy, et al. experiment rats were allowed to explore the choice-point region of a T maze; they could see into the two goal arms, but were prevented from entering them by means of glass doors. During the exposure period, both arms were similar in brightness, e.g., black; prior to a subsequent choice trial one of the arms was changed in brightness, e.g., to white. The rats tended to enter the arm which had been changed in brightness.

This experiment, incidently, makes of sI_R , Hull's *conditioned inhibition*, (17) an unlikely explanation of alternation, something which the Montgomery (25) and Glanzer (13) + maze studies do not accomplish. For, if the R of sI_R can refer to an *approach response*, conditioned inhibition remains a possible Hullian explanation of the + maze results, even though

I_R is not applicable. But in the Kivy, Earl, and Walker study, any approach response was minimal since the goal arms were blocked at the choice point by glass doors.

Both I_R and sI_R as well as sI are considered unsatisfactory explanations of the results of an experiment by Dember (2). In this study the procedure was similar to that followed by Kivy, et al., except for one important difference. On the exposure trial, one arm was black and the other white; on the choice-trial the arms were either both black or both white. The Ss were faced with a choice between two equally "satiated" or "inhibiting" stimuli. Their responses should, therefore, have been randomly distributed, but they were not: the Ss entered the arm which had been changed from the exposure-trial condition.

Dember's explanation of this behavior, and of alternation in general, makes use of the concept of *environmental change*, which is one source of *novelty*. This idea is further elaborated by Dember and Earl (3).

As far as alternation behavior is concerned, there is little to distinguish the Dember and Earl explanation from the one proposed by Montgomery (23, 24, 25). It is likely, however, that differences will emerge as the general theories from which these explanations are derived become further articulated.

Similarity of the alternatives. In accordance with sI deduction (g) the more similar the alternatives, the less should be the amount of alternation.⁴

Two direct tests of this hypothesis have been attempted. In the first

³ Unfortunately for this argument, Sutherland's rats were also confined in a goal-box away from the choice point.

⁴ It is interesting that Saltz (32) derives the opposite prediction from a theory which bears indirectly on alternation.

study⁵ alternation behavior under three similarity conditions was observed: rats were run in a **T** maze with (a) one black and one white arm, (b) one arm black and the other light grey, and (c) one black, the other dark grey. No differences in amount of alternation were found.

The second experiment relates to the similarity issue through Glanzer's further deduction that "the greater the extent of *S*'s sense-organ damage, the less the amount of spontaneous alternation" (16, p. 263). To test this hypothesis, the alternation behavior of a group of blinded rats was compared with that of a normal group; again no differences were found (5).

These failures to confirm predictions based on the similarity issue, as pointed out by Dember and Roberts (5), are most likely attributable to the rats' using dimensions other than vision as sources of alternation when visual differences are minimal or lacking. This suggests that any test of the similarity hypothesis must employ methods less direct than the ones used above. In particular, dimensions other than the one manipulated must be rendered neutral with respect to the two alternatives. Supporting evidence for the similarity hypothesis has been obtained through such indirect tests.

One test has been reported by Dember and Millbrook (4). In this experiment rats were first exposed, in the choice-point region, to the two arms of a **Y** maze. On the exposure trial one arm was grey and the other either black or white. On the choice trial, the two arms were equal in brightness, and both different in

brightness from the values used on the exposure trial. This means that the rat was faced with a choice between two changes, one greater than the other, as, for example, a black-to-white change versus a grey-to-white change. The *Ss*' behavior confirmed the deduction that the larger change, or the more dissimilar alternative, would be preferred.⁶

Another indirect confirmation of the similarity prediction is contained in the previously mentioned experiment of Walker, et al. (41), where the two responses were made more discriminable than usual. This was accomplished by means of a specially designed **+** maze which required a three dimensional response: to go right, the rat had to turn, twist, and climb right, and to go left, do just the opposite. Under these conditions, a significant, though still small amount of *response alternation* was obtained, whereas, in an otherwise comparable flat maze response alternation was at a chance level.

Also compatible with the similarity hypothesis are the results of an experiment by Zeaman and Angell (46). In this study, rats were forced to one arm of a four-arm radial maze. The forced trial was always to one of two arms placed at a 90° angle from the starting stem. The other three arms were separated from the forced arm by 60°, 120°, and 180°. After 10 forced, reinforced trials to the **T** arm, the rat was given a free trial with all four arms available. Frequency of choice was inversely related to angle of separation. If separation angle is thought of as contributing to similar-

⁵ E. L. Walker, unpublished manuscript, University of Michigan.

⁶ Dember and Millbrook suggest that this preference for the greater change may be useful as an indicator response for purposes of stimulus scaling.

ity, these results provide additional support for the similarity hypothesis.

An instructive failure to find a relation between amount of alternation and separation angle is provided in an experiment by Jackson (19). Rats were run in a maze with two arms separated by either 15°, 90°, or 180°. No differences were found in the amount of alternation produced by each condition. The same interpretation apparently applies here as was offered of the two direct tests described above. Given an opportunity to alternate, rats will do so as long as the alternatives are discriminable on some dimension. Decreasing the discriminability of the alternatives on one dimension, e.g., spatial separation, will not affect amount of alternation per se. The effect of these manipulations can be observed, however, in indirect tests, in which, for example, the *S* is allowed to choose among alternatives which vary in their similarity to the exposure stimuli.

Intertrial interval. Any theory of alternation must postulate some trace on Trial 2 of the events of Trial 1. The duration and strength of this trace are revealed by the relation between percentage of alternation and length of intertrial interval. Data of this sort are available, but quite varied.

For example, Montgomery (23) failed to obtain alternation for intervals greater than about a minute. Dennis (7) and Heathers (16) report significant alternation only for intervals up to about 2 min. Riley and Shapiro (30) found alternation with a 25 sec. interval, but not with one of 5 min.

Walker (38), however, has obtained significant amounts of alterna-

tion for much longer intertrial intervals. He related percentage of alternation to intertrial interval for intervals ranging from about one to 300 min. Amount of alternation varied around 75% for intervals up to 60 min. and then abruptly dropped to the chance level at about 90 min. This relationship between alternation and intertrial interval is not the decreasing negatively accelerated function expected either from the Hullian assumption about the dissipation of I_R or from Glanzer's analogous assumption relative to satiation. Heathers (16), however, did find an inverse relationship over very short intervals ranging from 15 sec. to 2 min.

There are plausible bases for explaining the wide discrepancies among these data. First, theoretical import has been ascribed to the degree of similarity between the alternatives. While it is true that direct tests have failed to reveal an effect of similarity on amount of alternation, these tests involved relatively short intertrial intervals. It seems likely that a more difficult test, i.e., one employing a long intertrial interval, would show an effect of similarity. Thus, experiments which differ with respect to the similarity variable might also be expected to yield different results pertaining to intertrial interval. To illustrate, Walker (38), employing a **T** maze with highly differentiated arms, found alternation with intervals up to at least an hour; Montgomery (23), who used highly similar alternatives, found no alternation with intervals beyond a minute.

The second factor to be considered in explaining the variety of intertrial interval results is the procedure

whereby the Ss experience the alternatives on Trial 1. The experiments above all used a *free-choice* procedure: both alternatives were available to the rat on Trial 1. With this, in contrast to a *forced-choice* procedure, the rat's typical vacillation at the choice point permits a partial experiencing of both alternatives; this experience may have the effect of reducing the alternation tendency on Trial 2.

Furthermore, the free-trial procedure invites the effect of bias, as discussed in the Section herein on Definition and Measurement. This would also lead to an apparent decrease in the amount of alternation if the bias effect were ignored in the treatment of the data.

With a forced-choice procedure only one alternative is available to the rat on Trial 1, thus decreasing both of the possible confounding effects mentioned above. A forced-choice procedure, therefore, should generally yield higher levels of alternation than a free-choice procedure.

Unfortunately, there is no single experiment in which intertrial interval has been varied following one forced trial. Nevertheless, it is possible to find cross-study data which support our hypothesis. Zeaman and House (47) obtained 61% alternation with a 60-min. intertrial interval between forced Trial 1 and free Trial 2. With the same procedure, but a 30-min. interval, Rothkopf and Zeaman (31) found alternation to be about 71%. Hence, significant alternation was obtained with the forced-trial procedure at intervals far exceeding all but those of Walker's (38) free-trial experiment.

In evaluating this set of data, it would seem reasonable to assume that positive results, i.e., alternation

after long intervals, reveal the basic relationship, and that negative results, i.e., no alternation after very brief intervals, are confounded by factors such as those we have cited. A serious experimental attack on this problem, in which degree of similarity of alternatives, Trial 1 procedure, and intertrial interval are varied, would be extremely valuable at this point.

Other experiments on the duration of alternation tendency might profitably be mentioned here. Under the condition of a free trial following 10 forced trials to the same arm of a T maze, Zeaman and House (47) were able to find alternation for intervals up to at least 12 hours. In studies where rats are trained to alternate, i.e., where reward on any trial is contingent on the S's alternating, Petrinovich and Bolles (29) and Ladieu (21) found that some rats could alternate after as much as a five hour interval. That trained alternation is differently motivated from spontaneous alternation is quite likely; nevertheless, both types presumably require similar trace mechanisms.

The postulation of a trace which decays over time suggests the possibility of using alternation as an indicator response for the study of short-term memory. This idea was proposed as early as 1939 by Dennis (7), but so far has received little attention. It has been followed in at least one experiment, however, which nicely illustrates such a use of alternation. Morgan and Wood (26) measured amount of alternation before and after making lesions in various cortical areas. They found a marked decrease in alternation scores following either frontal or occipital lesions.

Intratrial interval or exposure time.

In stimulus-oriented theories, the length of time during which the *S* is exposed to a stimulus object determines the amount of satiation or novelty-reduction produced, and hence, the amount of alternation. Glanzer (13) reports the only experiment directed at this assumption. Rats were detained for 10 minutes after their Trial 1 response in either of three places: the end box, the start box, or the choice-point region of a **T** maze. As predicted, the group detained in the end box showed the greatest amount of alternation, 98%, which was even higher than that of a no-delay group.

Distributions of successive free trials.

An interesting problem is posed when several free trials are given in succession at a fixed intertrial interval. Under these conditions it has generally been found that amount of alternation decreases from the first pair of trials to later pairs. Sutherland (35) gave his rats 11 trials per day with an intertrial interval of less than one minute. He obtained 81% alternation between Trials 1 and 2, but an average over all 10 pairs of only 65%. In another condition yielding a generally lower level of alternation the respective values were 65% and 49%.

Similar results were obtained earlier by Wingfield and Dennis (45). Over six massed trials, the amount of alternation between successive pairs decreased linearly from about 80% to 50%. In an experiment by Weitz and Wakeman (43), 30 pairs of trials were given; in successive blocks of 10 pairs of trials percentage of alternation was found to be 74, 60, and 53 for one condition, and 69, 48, and 48 for another. This same decline of alternation behavior over trials was clearly

obtained by Riley and Shapiro (30), but somewhat less clearly by Heathers (16) and Glanzer (13).

Though the data generally indicate a decrease in alternation over trials, irregularities and reversals in this trend are often present. A more definitive approach to this matter might be achieved through control of the *S*'s pattern of choices over pairs of trials, perhaps by the use of the forced-trial procedure.

One explanation of the decrease in alternation over trials appears obvious. On Trial 2 *S* is influenced only by the trace of the Trial 1 events, whereas on Trial 3 both the Trial 1 and Trial 2 traces may be active. If *S* has alternated between trials one and two, either alternative on Trial 3 will at least partially satisfy the alternation tendency. Glanzer (14) and Sutherland (35) have offered similar explanations.

A simple test of this explanation also seems obvious. The decrease in alternation over successive trials could be minimized under conditions of optimal trial spacing. This optimal condition would be realized if the interval between trials were short enough to yield a high percentage of alternation between Trials 1 and 2, but long enough so that by Trial 3 the trace from Trial 1 had dissipated. By inference from the intertrial data, an interval somewhere between 15 and 30 min. should meet these requirements.

Spatially successive alternatives. A response-oriented theory of alternation predicts the occurrence of alternation not only in the two-alternative, two-trial situation, but also in situations providing spatially successive choice points, as for example in a multiple **T** maze. This same predic-

tion would be made by a stimulus-oriented theory which takes cognizance of response-feedback as a potential source of alternation; response alternation would not be expected, however, as long as external stimulus differences are sufficiently salient.

The only clear evidence for successive response alternation comes from an experiment by Dennis and Henneman (8), in which a multiple **T** maze, composed of highly similar elements, was used. In a subsequent study employing a double square maze, Dennis (7) found that the response at the first choice point had no effect on that at the second choice point. But, on the rats' second run in the maze the response previously made at each choice point was alternated.

It may be that the lack of successive response alternation in the Dennis (7) experiment is attributable to the shape of the maze units. Before arriving at the second choice point, a rat would have made not one, but rather, four prior turns—two rights and two lefts. This should mask any response alternation tendency.

In a recent set of experiments, Estes and Schoeffler (12) used a forced-trial procedure with several variations of a multiple **T** maze. They, like Dennis, found that the "alternation tendency produced by a forced turn is almost entirely specific to the point in the maze at which the forced turn occurs" (12, p. 359). Of course, Estes and Schoeffler, unlike Dennis and Henneman, were pitting response alternation against stimulus alternation, and the predominance of the latter is not surprising.

ALTERNATION BEHAVIOR AND LEARNING THEORY

Alternation behavior, as a psycho-

logical issue, makes closest contact with learning theory in those cases where alternation occurs despite reinforcement of the first trial response. As Estes and Schoeffler state it: "According to any extant version of S-R theory, the reinforcement of runs to a given side of a **T** maze should increase the tendency to go to that side . . ." (12, p. 357). Clearly this is not the case.

The failure of reward to increase the probability of the occurrence of a response in the two-alternative situation is cogently demonstrated by studies in which a number of forced, reinforced trials are given to one arm of a **T** maze. Even with 10 such trials, alternation on a subsequent free-choice trial is very markedly present (47).

In fact, alternation tendency appears to be strengthened as the number of forced, reinforced trials to one alternative is increased. Both Rothkopf and Zeaman (31) and Zeaman and Angell (46) found more alternation after 10 forced trials than after one and two, respectively. In the study by Zeaman and House (47) the number of forced, reinforced trials to one side of a **T** maze was varied over a range from one to ten. Their data indicate a direct relation between alternation tendency and the number of reinforced trials to the forced alternative.

An interesting variation of the forced-trial technique has been employed by Denny (10). Rats were given two trials a day for 24 days in a **T** maze, with forced trials introduced in such a way that one arm was entered twice as often as the other. All trials were reinforced. During this training period there was an increasing tendency for the Ss to choose the less frequently entered (and rewarded) arm when allowed a free-

trial. On the day following the training period, two free trials were given, and again a significant preference for the less frequently entered arm was evidenced. The same was still true when two free trials were given one week later.

A striking aspect of these data concerns the Ss' behavior on the second free-trial: with respect to the first free-trial response, the second was a repetition, not an alternation. With respect to the entire sequence of previous responses, of course, the second free-trial response, as well as the first, was to the less frequently visited arm. In this sense, these data fit better with a novelty than a temporary satiation explanation of alternation.

The most direct attempt at handling this reinforcement "paradox" is Walker's (38) system, which was considered in the section herein on General Explanations of Alternation. In it is presented the view that reward can be conceived of as an "emphasizer." More specifically: "Every reaction produces a reaction decrement—a temporary lowering of the probability of the elicitation of the reaction. Reward contiguous with the reaction produces a reinforcement of the reaction. The more reinforcement there is, the more reaction decrement there will be. The decrement is followed in time by an increment. The more reinforcement there is, the more rapid the recovery from the decrement and the greater the eventual increment (learning)" (38, p. 167, italics omitted).

To test this idea, Walker compared the alternation tendencies of reinforced and nonreinforced rats over a lengthy range of intertrial intervals. Some evidence was obtained to support the hypothesis that reinforcement enhances alternation behavior

during the course of the reaction decrement.

The results of a recent unpublished study by Fowler and Fowler,⁷ however, are opposed to Walker's hypothesis. Rats were given six massed free trials in a T maze. Reinforcement consisted in the complete reduction in both goal arms of various electric shock intensities present in the starting stem. The data revealed an inverse relation between alternation tendency and magnitude of shock reduction. Moreover, the highly rewarded Ss showed not a single occurrence of alternation.

Related to this finding are the results of a study by DeValois (11) in which different kinds and degrees of motivation were related to variability of behavior. Employing two levels of thirst and of shock motivation, DeValois found amount of variability in a spatially successive, multichoice situation to be an inverse function of amount of motivation. These data, and more directly those of Fowler and Fowler, appear to indicate that alternation behavior is decreased, or eliminated, under conditions of high motivation and/or large magnitudes of reinforcement. It also is conceivable that escape or avoidance behavior generally shows less alternation than approach behavior, independently to strength of motivation and amount of reinforcement.

At any rate, as far as moderate drive and reinforcement are concerned, the problem posed by Estes and Schoeffler persists. For example, under the unequal reinforcement conditions of the Denny (10) study, the Ss tended to choose the less frequently rewarded arm. Even where

⁷ Fowler, H. and Fowler, D. E., unpublished manuscript, Yale University.

only one approach response is reinforced, as in the selective learning situation, alternation is usually prevalent in the early training trials.

From the point of view of the learning theorist, then, alternation behavior presents a serious problem. An equally difficult problem exists for the alternation theorist. "Why reinforcement disrupts the alternation pattern" is, for the latter, as crucial an issue as "why reinforced responses are alternated" is for the former. The solution of either version of the problem should contribute to the enrichment of behavior theory in general.

SUMMARY AND CONCLUSIONS

The spontaneous alternation behavior exhibited by rats faced with repeated choices of alternatives can no longer be explained adequately via Hull's concept of reactive inhibition. The source of alternation may sometimes be in the Ss' responses, but more generally alternation is a reaction to stimuli, of which response feedback is but one minor component. To replace the reactive inhibition explanation of alternation an analogous concept, stimulus satiation, has been offered by Glanzer. A theory built around this concept has received empirical support from a variety of experiments. Some data, however, seem to require a more general theoretical explanation, and motivational concepts, such as curiosity or response to novelty, have been suggested.

Certain empirical generalizations about the variables which affect alternation seem warranted: (a) If the alternatives require equal amounts of work, then variation in this amount of work does not affect alternation. (b) Variation in the

similarity of the alternatives on a single dimension does not influence amount of alternation, at least for short intervals; Ss, however, will choose from among alternatives the one which is the least similar to a previously experienced alternative. (c) Alternation may occur, following a single response, after an intertrial interval of as much as 90 min.; as number of forced responses to a single alternative increases, amount of alternation on a subsequent free-trial increases, reaching a maximum after 10 forced trials. With 10 forced trials alternation may occur after an intertrial interval as great as 12 hours. (d) Increasing the time that S is in contact with an alternative increases the amount of alternation. (e) As number of massed free trials increases, amount of alternation decreases. (f) If S is presented with a set of spatially successive choice points, the amount of alternation from one to the next is small, if above chance, and probably represents the relatively weak influence of response feedback as a source of alternation. (g) Under moderate drive strength and with moderate amounts of reinforcement, equal reinforcement in both alternatives does not decrease amount of alternation, as compared with a nonreinforcement condition. Even unequal reinforcement between the alternatives yields a high percentage of alternation. (h) Under conditions of strong drive and/or large magnitudes of reinforcement, potentially equal reinforcement between alternatives yields a low percentage of alternation.

The evidence which bears on the above generalizations has been examined, and in several places suggestions for further research have been

made. Attention was called to the equally difficult and equally important problems which the alternation-

reinforcement relation poses for the learning theorist on the one hand and the alternation theorist on the other.

REFERENCES

1. CRANNELL, C. W. Hesitation time and the solution of an alternation problem in rats. *J. Psychol.*, 1940, **9**, 379-385.
2. DEMBER, W. N. Response by the rat to environmental change. *J. comp. physiol. Psychol.*, 1956, **49**, 93-95.
3. DEMBER, W. N., & EARL, R. W. Analysis of exploratory, manipulatory, and curiosity behavior. *Psychol. Rev.*, 1957, **64**, 91-96.
4. DEMBER, W. N., & MILLBROOK, B. A. Free-choice by the rat of the greater of two brightness changes. *Psychol. Rep.*, 1956, **2**, 465-467.
5. DEMBER, W. N., & ROBERTS, W. Alternation behavior in peripherally blinded rats. *Percept. Mot. Skills*, in press.
6. DENNIS, W. A comparison of the rat's first and second explorations of a maze unit. *Amer. J. Psychol.*, 1935, **47**, 488-490.
7. DENNIS, W. J. Spontaneous alternation in rats as an indicator of the persistence of stimulus traces. *J. comp. Psychol.*, 1939, **28**, 305-312.
8. DENNIS, W., & HENNEMAN, R. H. The non-random character of initial maze behavior. *J. genet. Psychol.*, 1932, **40**, 396-405.
9. DENNIS, W., & SOLLENBERGER, R. T. Negative adaption in the maze exploration of rats. *J. comp. Psychol.*, 1934, **18**, 197-205.
10. DENNY, M. R. Learning through stimulus satiation. *J. exp. Psychol.*, 1957, **54**, 62-64.
11. DEVALOIS, R. C. The relation of different levels and kinds of motivation to variability of behavior. *J. exp. Psychol.*, 1954, **47**, 392-398.
12. ESTES, W. K., & SCHOEFFLER, M. S. Analysis of variables influencing alternation after forced trials. *J. comp. physiol. Psychol.*, 1955, **48**, 357-362.
13. GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387-393.
14. GLANZER, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257-268.
15. GLANZER, M. Stimulus satiation in situations without choice. Paper read at Eastern Psychological Association, New York, April, 1957.
16. HEATHERS, G. L. The avoidance of repetition of a maze reaction as a function of the time between trials. *J. Psychol.*, 1940, **10**, 359-380.
17. HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
18. IRWIN, F. W., & PRESTON, M. G. Avoidance of repetition of judgments across sense modalities. *J. exp. Psychol.*, 1937, **21**, 511-520.
19. JACKSON, M. M. Reactive tendencies in the white rat in running and jumping situations. *J. comp. physiol. Psychol.*, 1941, **31**, 255-262.
20. KIVY, P. N., EARL, R. W., & WALKER, E. L. Stimulus context and satiation. *J. comp. physiol. Psychol.*, 1956, **49**, 90-92.
21. LADIEU, G. The effect of length of delay interval upon delayed alternation in the albino rat. *J. comp. physiol. Psychol.*, 1944, **37**, 273-286.
22. LEPEV, W. M., & RICE, G. E., JR. Behavior variability in paramecia as a function of guided act sequences. *J. comp. physiol. Psychol.*, 1952, **45**, 283-286.
23. MONTGOMERY, K. C. Spontaneous alternation as a function of time between trials and amount of work. *J. exp. Psychol.*, 1951, **42**, 82-93.
24. MONTGOMERY, K. C. Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.*, 1952, **45**, 50-57.
25. MONTGOMERY, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45**, 287-293.
26. MORGAN, C. T., & WOOD, W. M. Cortical localization of symbolic processes in the rat. II. Effect of cortical lesions

- upon delayed alternation in the rat. *J. Neurophysiol.*, 1943, **6**, 173-180.
27. MOWRER, O. H. & JONES, H. M. Extinction and behavior variability as functions of effortfulness of task. *J. exp. Psychol.*, 1943, **33**, 369-386.
 28. PAVLOV, I. P. *Conditioned reflexes*. (G. V. Anrep, Trans.) London: Oxford Univer. Press, 1927.
 29. PETRINOVICH, L., & BOLLES, R. Delayed alternation: Evidence for symbolic processes in the rat. *J. comp. physiol. Psychol.*, 1957, **50**, 363-365.
 30. RILEY, D. A., & SHAPIRO, A. N. Alternation behavior as a function of effortfulness of task and distribution of trials. *J. comp. physiol. Psychol.*, 1952, **45**, 468-475.
 31. ROTHKOFF, E. Z., & ZEAMAN, D. Some stimulus controls of alternation behavior. *J. Psychol.*, 1952, **34**, 235-255.
 32. SALTZ, E. A single theory for reminiscence, act regression, and other phenomena. *Psychol. Rev.*, 1953, **60**, 159-171.
 33. SIEGEL, P. S. Reactive inhibition as a function of number of response evocations. *J. exp. Psychol.*, 1950, **40**, 604-608.
 34. SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1-40.
 35. SUTHERLAND, N. S. Spontaneous alternation and stimulus avoidance. *J. comp. physiol. Psychol.*, 1957, **50**, 358-362.
 36. TELFORD, C. W. The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1-36.
 37. TOLMAN, E. C. Purpose and cognition: The determiners of animal learning. *Psychol. Rev.*, 1925, **32**, 285-297.
 38. WALKER, E. L. The duration and course of the reaction decrement and the influence of reward. *J. comp. physiol. Psychol.*, 1956, **49**, 167-176.
 39. WALKER, E. L., DEMBER, W. N., EARL, R. W., & KAROLY, A. J. Choice alternation: I. Stimulus vs. place vs. response. *J. comp. physiol. Psychol.*, 1955, **48**, 19-23.
 40. WALKER, E. L., DEMBER, W. N., EARL, R. W., FLIEGE, S. W., & KAROLY, A. J. Choice alternation: II. Exposure to stimulus or stimulus and place without choice. *J. comp. physiol. Psychol.*, 1955, **48**, 24-28.
 41. WALKER, E. L., DEMBER, W. N., EARL, R. W., FAWL, C. L., & KAROLY, A. J. Choice alternation: III. Response intensity vs. response discriminability. *J. comp. physiol. Psychol.*, 1955, **48**, 80-85.
 42. WATANABE, M., & IWATA, S. Alternative turning response of *Armadillidium Vulgare*. *Annu. anim. Psychol.*, 1956, **6**, 75-82.
 43. WEITZ, J., & WAKEMAN, M. L. "Spontaneous" alternation and the conditioned response. *J. comp. Psychol.*, 1941, **32**, 551-562.
 44. WINGFIELD, R. C. Some factors influencing spontaneous alternation in human subjects. *J. comp. physiol. Psychol.*, 1943, **35**, 273-243.
 45. WINGFIELD, R. C., & DENNIS, W. The dependence of the rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, **18**, 135-147.
 46. ZEAMAN, D., & ANGELL, D. A spatial gradient of alternation tendency. *J. comp. physiol. Psychol.*, 1953, **46**, 390-392.
 47. ZEAMAN, D., & HOUSE, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41**, 177-186.

Received March 24, 1958.

THE CONTINUITY OF ABNORMAL AND NORMAL BEHAVIOR

H. J. EYSENCK

Institute of Psychiatry, University of London

In a recent paper in this journal, Pearson and Kley (11) complain of

the proclivity of psychologists for assuming or demonstrating variables to be distributed in continuous fashion throughout the general population, but concentrating attention on the pathological extremes, which may, in fact, constitute discrete series. The familiar concept of a normal distribution for "emotional adjustment" ranging from "super-normal" and "normal" to "neurotic" and "psychotic" has led laymen and behavioral scientists alike to picture human emotions in various shades of gray. . . . While such conceptualizations may serve a useful purpose, they may also be misleading. The danger lies in the temptation to infer continuous distribution of underlying etiological factors from the fact that behavioral traits appear to be so distributed.

Pearson and Kley then quote a study by Eysenck and Prell (8) as an apparent example of this fallacy. They say of these authors that:

their assumption that neuroticism is on a continuum in the general population and the samples employed make (*sic*) it impossible to infer that clinical cases of neurosis arise at the extreme end of the continuum only because of the *degree* to which they inherit the neuroticism factor. Testwise or symptomwise, the diagnosed neurotics do constitute the extreme of the distribution, but the reason for their coming to this sorry end may be quite different from the reasons which cause individuals in the "borderline" or "normal" range of tests scores or clinical behavior to fall where they do.

The point that the distribution of scores on a single test cannot be safely interpreted to give a correct indication of the distribution of the underlying determinants in the absence of a proper metric and in view of the usual large error variance is well taken. It was made explicitly by the

writer in *The Structure of Human Personality* (5, p. 11), and having always argued against the tendency of psychiatrists and psychologists to *assume* either continuity or discontinuity of normal and abnormal behaviour in the absence of *proof*, the writer was not unnaturally surprised to find himself accused of this very crime. It is the purpose of this brief note to show that this very fundamental criticism of the genetic and other experimental work done in the field of abnormality by the writer and his colleagues is not in fact subject to this charge.

We have already agreed with Pearson and Kley that no faith can be put in the distribution of scores on any one test in arguing for or against the continuity hypothesis. The writer accordingly put forward a method for investigating this problem along hypothetico-deductive lines which was published under the name of "Criterion Analysis" (2). Having outlined his theory of the existence of the general factor of neuroticism, similar in mode of derivation and general interpretation on the orectic side to the general factor of *intelligence* on the cognitive side, the writer went on to say that what was at issue in this paper was "the hypothesis that this putative factor of 'neuroticism' forms a quantitative continuum, on one extreme of which are to be found hospitalized neurotics, while so-called normals are to be found all the way from the near neurotic and neurotic to the conspicuously non-neurotic, mature, stable,

and integrated type of personality" (2, p. 42). It was also pointed out that a similar problem arose in connection with Kretschmer's hypothesis of the existence of a normality-abnormality continuum ranging from the normal to the psychotic. Having thus indicated the problem, which of course also includes the identity or independence from each other of these two hypothetical continua, the writer went on to outline the method of criterion analysis, whose specific merit was claimed to be its ability to provide evidence relevant to this type of hypothesis. Empirical data have been given to demonstrate the truth of the continuity hypothesis, particularly with respect to neuroticism (2) and psychoticism (3). Later studies using the technique of canonical variate analysis (6, 9) have demonstrated the essential independence of these two continua. A detailed discussion of all this work is given in *The Dynamics of Anxiety and Hysteria* (7).

It would be open to Pearson and Kley to criticise this method along various lines. They might argue against the logic underlying criterion analysis which postulates that if, and only if, there exists a continuum between normal and abnormal mental states will there be found (a) corresponding factors in correlation matrices derived separately from tests administered to normal and abnormal groups and (b) significant correlations between these factor loadings and what the writer has called the "criterion column," i.e., the column of biserial correlations between normality-abnormality on the one hand, and the various tests used in the experiment on the other. Such attempts as have been made in the literature to impugn the logical validity

of criterion analysis (e.g., 1) appear to have rested on a misunderstanding of the method (4) and cannot be regarded as fatal to the postulation.

Secondly, it might be open to Pearson and Kley to argue that the specifications of the method are not sufficiently clearly expressed to make its use feasible. Thus a large and varied battery of tests is required, all of which must discriminate significantly between the normal and the abnormal group. It might be argued that "large and varied" is too indefinite a description, and that no particular standard of significance has been specified. These objections would be well taken, and it is to be hoped that in due course it will prove possible to give a more operational definition of "varied" than is possible at present. The writer doubts, however, if these difficulties are fatal to the method. Until better criteria of selection are available, it might be suggested that the number of tests should not be below 20, with the standard of significance to be taken as the $p=.01$ level, and that the term "varied" should be interpreted as referring to the abilities involved in the tests used, as determined by factorial analysis; different muscle groups used in the execution of the tasks; different sense organs used in the mediation of the tasks; and so forth. It is doubtful whether in practice much doubt will arise on any of these points.

Lastly, it will be open to the critic to point to certain weaknesses in the mathematical treatment of criterion analysis, as was done for instance by Lubin (10). These difficulties are very real, and the writer has no wish to gloss over them. Until they are completely overcome it is obviously necessary to use the method with

considerable circumspection, and only with the fullest understanding of the assumptions underlying each step taken. Nor should the interpretation of the final results be made in any but the most tentative fashion. Nevertheless, and in spite of all these qualifications, the writer does not know of any other method available at present which tackles this particular problem, or which can offer us worthwhile information relating to it. Until a better method is available, therefore, criterion analysis will remain as a worthwhile addition to our methodological set of tools.

It is noteworthy, however, that Pearson and Kley do not criticise criterion analysis on any of these grounds. What they do instead is to neglect the whole body of work done by the writer in connection with this method and this problem and to present him as basing his views entirely on an invalid argument from the simple distribution of single test scores. This does not appear to the writer to be a reasonable form of criticism, and consequently it seems desirable to put the point in its proper perspective.

In spite of all that has been said in this note, the writer would not wish to dismiss the possibility or even the likelihood that in any random group of clinically diagnosed neurotics there would be found a small number of people who might "constitute a group apart, different not in degree, but in kind, by reason of some specified bio-

chemical error, which is highly predictable in terms of inheritance, and which operates in a manner quite different from anything observed" in the kinship relations of the remainder of that group. The evidence quoted makes it somewhat unlikely that the major part of any given neurotic or psychotic group would be made up of such individuals, but no one familiar with the heterogeneity of psychiatric groups would wish to suggest seriously that all the members of such groups were homogeneous with respect to hereditary processes and genetic determinants. Nevertheless, *as a first approximation*, we must be concerned with the major sources of variance affecting the majority of members of such groups, and it is in this connection that the writer cannot follow the criticism levelled by Pearson and Kley against the Eysenck and Prell study.

SUMMARY

Pearson and Kley (11) criticize the writer for basing his belief in the continuity of normal and abnormal states on the invalid consideration that test scores tended to be continuous between the groups. In answer, the writer has pointed out that he himself had discussed the lack of validity of this procedure in detail and had advocated a different method, namely, that of criterion analysis, specifically designed by him to deal with problems of this kind.

REFERENCES

1. BEEZHOLD, F. W. On criterion analysis. *J. Nat. Inst. Personnel Res.*, 1953, 5, 176-182.
2. EYSENCK, H. J. Criterion analysis—an application of the hypothetico-deductive method to factor analysis. *Psychol. Rev.*, 1950, 57, 38-53.
3. EYSENCK, H. J. Schizothymia-cyclothymia as a dimension of personality. *J. Personal.*, 1952, 20, 345-384.

4. EYSENCK, H. J. On criterion analysis. A reply to F. W. Beezhold. *J. Nat. Inst. Personnel Res.*, 1953, 5, 183-187.
5. EYSENCK, H. J. *The structure of human personality*. London: Methuen, 1953.
6. EYSENCK, H. J. Psychiatric diagnosis as a psychological and statistical problem. *Psychol. Rep.*, 1955, 1, 3-17.
7. EYSENCK, H. J. *The Dynamics of Anxiety and Hysteria*. London: Routledge and Kegan Paul, 1957.
8. EYSENCK, H. J., & PRELL, D. The inheritance of neuroticism: An experimental study. *J. ment. Sci.*, 1951, 97, 441-465.
9. EYSENCK, S. B. G. Neurosis and psychosis as experimental analysis. *J. ment. Sci.*, 1956, 102, 517-529.
10. LUBIN, A. A note on "criterion analysis." *Psychol. Rev.*, 1950, 57, 54-57.
11. PEARSON, J. S., & KLEY, I. B. On the application of genetic expectancies as age-specific base rates in the study of human behavior disorders. *Psychol. Bull.*, 1957, 54, 406-420.

Received March 10, 1958.

DISCONTINUITY AND CORRELATION: A REPLY TO EYSENCK

JOHN S. PEARSON AND IRENE B. KLEY

Mayo Clinic, Rochester, Minnesota

Reference in our paper (3) to the study by Eysenck and Prell (1) was not intended to imply that they unwittingly subscribed to the error of inferring continuous distribution of etiologic factors giving rise to membership in either of the classes "neurotic" or "normal" from the fact that, phenotypically, the behavior of both classes seems to be on a continuum (2). Rather, our citation and discussion were intended to underscore the warning which Eysenck has duly set forth as to the heuristic nature of criterion analysis and the fact that this method and its application involve several arguable mathematical and logical assumptions. Also, it should be noted that our emphasis on this warning applies not merely to questionable inferences drawn from the distribution of scores on a single test but also to inferences drawn from a distribution, however complex and polydimensional its conception. The statement still stands that Eysenck and Prell's criterion group of "neurotics" may have contained some individuals so categorized by reason of *qualitatively* different determinants from those which determined the class membership of other "neurotics" and of the "normals." There is simply no means inherent in the experiment they have described by which the truth or falsity of this statement may be determined. Knowledge of the relative numbers of clinically diagnosed neurotics with qualitatively different determinants as against those who differ from normals only in the quantitative sense is of considerable importance to the ap-

plication of correlational measures in any effort to reveal the nature of underlying genotypic or causal factors. While the evidence we cited may not be particularly convincing as to the probable existence among clinical neurotics of a significant number of cases determined by a specific gene or gene pair, nevertheless, in the absence of any proof to the contrary, the possibility should be entertained as real. If this possibility is an actuality, a potentially serious inconsistency is introduced repeatedly in criterion analysis in the imposition of the assumption of continuity upon data where this assumption is not justified. This would be particularly true when computing biserial correlations between normality-abnormality and various measures of behavior. Whether or not this inconsistency is so trivial as to be safely ignored cannot be determined unless the method is proved to be adequate in an experiment where discrete primary determinants of membership in the phenotypic class, "abnormal," are firmly established on external grounds. The fact that opportunities for such experiments are rare in domains of human behavior germane to psychometrics has made it difficult for factor analysts to establish the validity of their various methods in the identification of causal factors.

The availability of one such opportunity was hinted in our paper and it was hoped that this might suggest a fruitful area of study to various workers. It may be well to amplify this suggestion briefly here. One might define operationally the phenotypic

classes "abnormal motor coordination" and "normal motor coordination" and select subjects whose membership in the former could be ascribed primarily to real, discrete, causative factors such as the genes for Huntington's chorea, Wilson's disease, Friedreich's cerebellar ataxia, familial spastic paralysis, and hereditary muscular dystrophy. To these might be added cases in which impaired motor coordination was observed to ensue following exposure to specific agents of trauma, infection, or toxemia (e.g. gunshot wound in the parietal area, mumps encephalitis, or lead encephalopathy). With samples sufficiently large in each category and a battery of motor tests as large and varied as the ingenuity of the experimenter could provide, one could subject a matrix of test intercorrelations to factor analytic procedures in the hope that these could work backward from the valid discrimination of the broad phenotypic categories "normal motor coordination" and "abnormal motor coordination" to the valid discrimination of the separate genotypes identified previously on bases external to the experiment. For this to be possible, it would be necessary for the mathematical treatment of the test variables to thread through a complex of intervening variables and hypothetical constructs between the known membership in one or another of the discrete genotypic classes and the observed membership in either the "normal" or "abnormal" classes along the phenotypic continuum of "motor coordination." Such complicating factors would include, for example, the biochemical action by which an individual gene mediates the reaction which we observe as Huntington's chorea and the second-

ary determinants of "normal" versus "abnormal" motor coordination such as a reactive depression or an overcompensatory reaction to awareness of progressive organic disability. The biggest obstacle would probably be the existence of interaction effects between the identified genotypes and all other genotypic determinants of membership in all other classes subsumed under all other behavior domains.

The reader may be impressed by the strictures which such an experiment would place upon factor analytic methods. Admittedly, negative results (i.e., failure to emerge with a factor structure closely analogous to the predetermined genotypic composition of the "abnormal" group) would not vitiate the heuristic application of similar procedures when there is no good reason to assume discrete distribution of causative factors. On the other hand, positive results in the experiment we have outlined would be an indication that the imposition of assumptions of continuity on data where these are known or suspected to be unjustified is not necessarily an inconsistency fatal to the method of criterion analysis. Experiments of this kind should be attempted by persons who are in a position to do them. At the same time, it would be well if someone were to come to grips with the formidable task of formalizing mathematical models on various discontinuity and interaction assumptions.

Our original point was to underscore Eysenck's own carefully stated reservations concerning the application of criterion analysis. We erred in not making this sufficiently clear. Here, we have tried to rectify this error and to amplify a suggestion made implicitly in the earlier paper,

for an experimental approach to the validation of factor analytic procedures in relation to discrete genotypic determinants of human behavior.

REFERENCES

1. EYSENCK, H., & PRELL, D. The inheritance of neuroticism: An experimental study. *J. ment. sci.*, 1951, **97**, 441-465.
2. EYSENCK, H. J. The continuity of abnormal and normal behavior. *Psychol. Bull.*, 1958, **55**, 429-432.
3. PEARSON, J., & KLEY, IRENE. On the application of genetic expectancies as age-specific base rates in the study of human behavior disorders. *Psychol. Bull.*, 1957, **54**, 406-420.

Received March 31, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY

This journal is now published by the American Psychological Association. It is a bimonthly; issues appear in February, April, June, August, October, and December. Contents include articles on problems of teaching, learning, and the measurement of psychological development.

All back issues and subscriptions up to and including the May 1957 issue are the property of Warwick and York, Inc., 10 East Centre Street, Baltimore 2, Maryland.

Subscriptions \$8.00
(Foreign \$8.50)

Single
Copies \$1.50

Direct new subscriptions and renewals to:

AMERICAN PSYCHOLOGICAL ASSOCIATION

Publications Office
1333 Sixteenth Street, N.W.
Washington 6, D.C.

Two new important evergreen Psychology Books

In fine soft-cover editions at \$1.35 and up

THE SEARCH WITHIN, by Theodor Reik. The famous psychoanalyst focuses upon himself with unsparring candor in this brilliant distillation of his thoughts and experiences. \$1.35

WHO ARE THE GUILTY? by David Ahlhammen, M.D. The burning question of juvenile delinquency receives its most penetrating answer to date as the famous psychiatrist details his clear-cut prescription for preventative action. \$1.35

Sexualities in Modern Man by Theodor Reik. \$1.35

A Study of Interpersonal Relations, Edited by Patrick Mullaney. \$2.45

Psychoanalytic Evolution & Development by Clara Thompson and Patrick Mullaney. \$1.75

Black Anger, A unique psychoanalytic biography of a native African medicine man, by Wolf Sachs. \$1.75

Sexuality and Human Relations by Ashley Montagu. \$1.45

The Forgotten Language, An Introduction to the Understanding of Dreams, Fairytales and Myths, by Erich Fromm. \$1.75

Rebel Without a Cause, The Psychoanalysis of a criminal psychopath, by Robert M. Loewner. \$1.75

The Blooded Nature of Man by Ashley Montagu. \$1.35

Democracy and Dictatorship, A study of the conditions that produce each, by Zeev Joel Barbu. \$1.45

Listening with the Third Ear by Theodor Reik. \$1.35

Anna's Psychology and Its Social Meaning by Leo Berg. \$1.75

Godless Myth and Creation by Patrick Mullaney. \$1.35

Specialized Techniques in Psychotherapy, Edited by Gustav Rothblatt, M.D., and J. Lewis Daport, M.D. \$1.45

The Idioms of Love, An essay by Sigmund Freud with selections from Dr. Reik's writings. \$1.75

At all bookstores • Send for free descriptive catalog GROVE PRESS, 77th Broadway, N. Y. 2

